

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**IN THE SPECIFICATION:**

**Please replace the following paragraph:**

**Paragraph beginning at page 1, line 3:**

This application is a continuation of U.S. Application No.09/370,295, filed August 9,1999 and now abandoned, which claims the benefit of U.S. Provisional Application No. 60/096,342, filed August 12, 1998.

**REMARKS**

In the Office Action Claims 13-15, 19-22 and 24 are rejected under U.S.C.112, first paragraph as the specification has not been found to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the invention commensurate in scope with these claims. Rejection of Claims 13-15, 19-22 and 24 is respectfully traversed on ground of the discussion below, which should also remove the examiner's objection to claims 16 and 17 on ground of their dependence on claim 13. A copy of Nosaka et al. (J. Biol. Chem. 268:17440-17447, 1993), enclosed for the Examiner's convenience (copy previously submitted with the IDS dated February 25, 2002), describes the isolation and characterization of a thiamin pyrophosphokinase gene from yeast (gi: 6324717). The expression of the yeast enzyme in *E. coli*, which lacks thiamin pyrophosphokinase activity, showed marked activity of this enzyme in the prokaryotic cell. Appendix A, enclosed herewith, shows a comparison of the claimed sequence with the yeast thiamin pyrophosphokinase and the mouse gene product (gi: 6468206), which share 18.7% and 24.3% sequence identity with the claimed sequence, respectively. All three sequences display essentially the same glycine-rich motif typical for nucleotide phosphate group binding and reminiscent of the Rossmann fold (Rossmann et al., Nature, Vol: 250, 194-199, 1974, copy enclosed for Examiner's convenience). Apart from thiamin, thiamin pyrophosphokinase also uses ATP as a substrate. Thus, the existence of such a binding site would be expected. Recently the crystal structure of the mouse enzyme has been revealed, which provides further support for above discussion (Timm et al., JMB, Vol:310, 195-201, 2001, copy enclosed for Examiner's convenience). The crystallographic data of the mouse enzyme identified residues of the thiamin pyrophosphokinase active site. All of the residues identified are essentially conserved among all three sequences. For the Examiner's convenience those residues have been underlined in the alignment in Appendix A.

This comparison demonstrates the sequence of the invention possesses stretches of highly conserved regions. One skilled in the art would appreciate that the

more highly conserved a residue is, the less likely that it could be modified and function maintained. From this alignment in the attached Appendix, one could quickly determine which amino acid residues might be modified in SEQ ID NO:2 without a likely change in function. Since SEQ ID NO:2 share 18.7% and 24.3% identity with the yeast and mouse protein, respectively, one of skill in the art would have appreciated that many variants sharing at least 80% sequence identity to the SEQ ID NO:2 would have been expected to retain Thiamin pyrophosphokinase activity.

In view of the discussion above, one skilled in the art would have known, at the time the claimed invention was made, how to use the claimed sequence without undue experimentation and that the rejection according to 35 USC §112, first paragraph, should be removed.

For the foregoing reasons, Applicants respectfully request reconsideration and allowance of the claims.

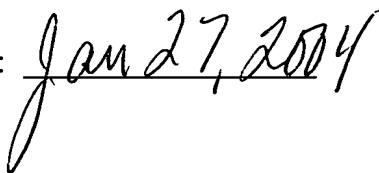
Please charge any requisite fee to Deposit Account No. 04-1928 (E. I. du Pont de Nemours and Company).

Respectfully submitted,



LORI Y. BEARDELL  
ATTORNEY FOR APPLICANTS  
REGISTRATION NO. 34,293  
TELEPHONE: (302) 992-4926  
FACSIMILE: (302) 892-1026

Dated: Jan 27, 2004





## APPENDIX A

Comparison of the amino acid sequences of the Thiamin Pyrophosphokinases from corn clone contig with SID No's: (SEQ ID NO:2), yeast and mouse set forth NCBI General Identifier No. 6324717 and 6468206, respectively. Amino acids conserved among all three sequences are indicated with an asterisk above the conserved residues, respectively. Dashes are used by the program to maximize alignment of the sequences. The likely nucleotide binding site (Rossmann et al., Nature, Vol: 250, 194-199, 1974) is boxed and residues involved in Thiamin binding (Timm et al., JMB, Vol:310, 195-201, 2001) are underlined.

```

*                                     *   ***
SEQ_id_2      MLWRAVRSM DVIMHSSSFL-----LPKLYQPVNKPVKNYALVVLNQHL P---RFMPRL
gi-6324717_yeast MSEECIENPERIKIGTDLINIRNKMNLKELIHPNED--ENSTLLILNQKIDIPRPLFYKI
gi-6468206_mouse ME-HAFTPLEPLLP TGNL-----KYCLVVLNQPLD---ARFRHL
```

```

*   *   **** *   *   *   *   *   *   *
SEQ_id_2      WDHANLRICADGGANHIFDEMYQITNDEDKKSTRNKYVPEIIIEGDMDSIRPEVKLFYSSQ
gi-6324717_yeast WKLHDLKVCADGAANRLYDYL-----DDDETLRIKYLPNYIIIGDLSLSEKVYKYRKN
gi-6468206_mouse WKKALLRACADGGANH---LYDLTEGE-----RESFLPEFVSGDFDSIRPEVKEYYTKK
```

```

*   **   *   *
SEQ_id_2      GSKISDKSHNQETTDLHKCISRIHHCTPDDEKPNLC-----
gi-6324717_yeast KVTII-KQTTQYSTDFTKCVNLISLHFNSPEFRSLISNKDNLQSNHGIELEKGIHTLYNT
gi-6468206_mouse GCDLISTP-DQDHTDFTKCLQVLQRKIEEKELQVDV-----
```

```

*   *****   *   *   *
SEQ_id_2      -----VLVTGALGGRFDHEAANINVLY-L---FSDMRIVLLSDDCLIRLLP
gi-6324717_yeast MTESLVFSKVTPISLLALGGIGGRFDQTVHSITQLYTLSENASYFKLCYMTPTDLIFLIK
gi-6468206_mouse -----IVTLGGLGGRFDQIMASVNTLF-QATHITPVPIIIIQKDSLIIYLLO
```

```

*   *   *   *   *   *   *   *   *   *   *
SEQ_id_2      RTHR----HELYIESSVEGPHCGLFVPGAPSTSTTTTGLKWNLSE-SKMRFGSMISTSNI
gi-6324717_yeast KNGTLIEYDPQFRNTCIGN--CGLLPIGEATLVKETRGLKWDVKNWPTS VVTGRVSSSNR
gi-6468206_mouse PG-K----HRLHVD TGMEGSWCGLIPV GQPCNQVTTTGLKWNLTN-DVLGFGTLVSTSN T
```

```

SEQ_id_2      VQSEK-VTVESDADLLWTISL-----RNLT
gi-6324717_yeast FVGDNCCFIDTKDDIILNVEIFVDKLIDFL
gi-6468206_mouse YDGSGLVTVETDHP L LWTMAI-----KS--
```

## Crystal Structure of Thiamin Pyrophosphokinase

David E. Timm\*, Jingyuan Liu, L.-J. Baker and Robert A. Harris

Department of Biochemistry  
and Molecular Biology, Indiana  
University School of Medicine  
Indianapolis, IN 46202, USA

Thiamin pyrophosphate (TPP) is a coenzyme derived from vitamin B<sub>1</sub> (thiamin). TPP synthesis in eukaryotes requires thiamin pyrophosphokinase (TPK), which catalyzes the transfer of a pyrophosphate group from ATP to thiamin. TPP is essential for central metabolic processes, including the formation of acetyl CoA from glucose and the Krebs cycle. Deficiencies in human thiamin metabolism result in beriberi and Wernicke encephalopathy. The crystal structure of mouse TPK was determined by multiwavelength anomalous diffraction at 2.4 Å resolution, and the structure of TPK complexed with thiamin has been refined at 1.9 Å resolution. The TPK polypeptide folds as an  $\alpha/\beta$ -domain and a  $\beta$ -sandwich domain, which share a central ten-stranded mixed  $\beta$ -sheet. TPK subunits associate as a dimer, and thiamin is bound in the dimer interface. Despite lacking apparent sequence homology with other proteins, the  $\alpha/\beta$ -domain resembles the Rossmann fold and is similar to other kinase structures, including another pyrophosphokinase and a thiamin biosynthetic enzyme. Comparison of mouse and yeast TPK structures reveals differences that could be exploited in developing species-specific inhibitors of potential use as antimicrobial agents.

© 2001 Academic Press

\*Corresponding author

Keywords: thiamin; crystal structure; vitamin; metabolism; kinase

### Introduction

Humans lack the ability to synthesize thiamin and, therefore, require dietary sources of this compound, also known as vitamin B<sub>1</sub>. Thiamin is converted to thiamin pyrophosphate (TPP) by thiamin pyrophosphokinase (TPK). TPP is essential for normal carbohydrate utilization as a coenzyme for central metabolic functions,<sup>1–3</sup> including the Krebs cycle as part of the  $\alpha$ -ketoglutarate dehydrogenase complex and the conversion of pyruvate to acetyl CoA catalyzed by the pyruvate dehydrogenase complex. The importance of thiamin in human and animal diets has been appreciated for a long time. Thiamin deficiency in humans causes beriberi and Wernicke encephalopathy. Beriberi has been known for centuries as a potentially fatal disease of the Far East resulting in neurologic and cardiovascular problems. Wernicke encephalopathy is a neurological condition often encountered in

chronic alcoholism due to reduced intake and uptake of dietary thiamin.<sup>4</sup> Inhibition of TPK by ethanol may also contribute to this condition.<sup>5,6</sup>

The formation of TPP is necessary for centrally important metabolic processes,<sup>1–3</sup> and the loss of TPK activity is lethal in yeast.<sup>7,8</sup> In addition to functioning as a carrier of activated acyl groups in  $\alpha$ -keto acid dehydrogenases required for glucose metabolism, TPP is required for carbohydrate metabolism by transketolase, pyruvate decarboxylase and pyruvate oxidase. TPP is also a coenzyme for the branched-chain  $\alpha$ -keto acid dehydrogenase complexes and acetolactate synthase required for branched chain amino acid degradation and synthesis,<sup>2</sup> respectively. TPP formation is also important in signaling and driving cellular transport of thiamin.<sup>2,8</sup> TPP levels regulate mitochondrial

$\alpha$ -keto acid dehydrogenases,<sup>9</sup> and feedback inhibition of TPK by TPP<sup>10–12</sup> may be a mechanism for maintaining proper TPP levels as a regulatory feature of these central metabolic enzymes. TPP also regulates transcriptional levels of thiamin biosynthetic enzymes in yeast and effects mating in fission yeast.<sup>2,8</sup> Finally, TPP is the precursor of thiamin triphosphate, which may have non-cofactor effects on neurologic cells.<sup>13</sup>

Abbreviations used: HPPK, hydroxymethylidihydropterin pyrophosphokinase; MAD, multiwavelength anomalous diffraction; PEG, polyethylene glycol; PRPP, phosphoribosylpyrophosphate synthetase; TPK, thiamin pyrophosphokinase; TPP, thiamin pyrophosphate.

E-mail address of the corresponding author: [dtimm@iupui.edu](mailto:dtimm@iupui.edu)

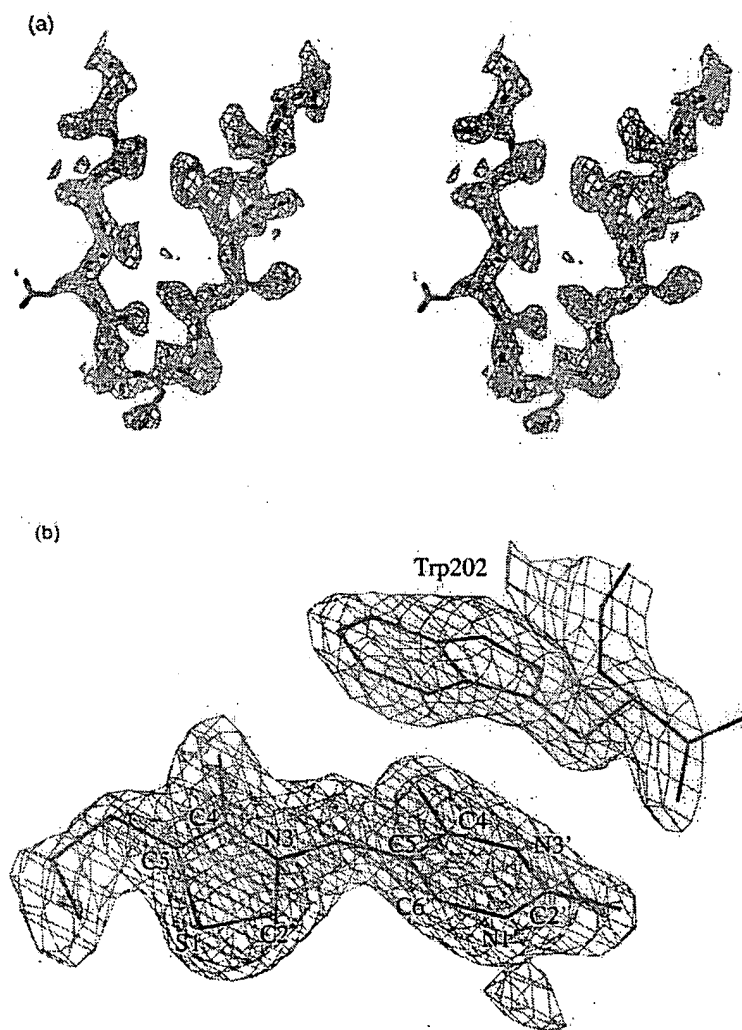
TPK has been characterized using material isolated from natural<sup>10-12,14</sup> and recombinant sources.<sup>7,8,15</sup> TPK amino acid sequences are poorly conserved between species, and database searches have revealed no sequence similarities to other kinases or other families of proteins. TPK from different species has subunit masses ranging from 23-36 kDa associated as dimers in solution.<sup>10,14</sup> TPK utilizes ATP and thiamin as substrates in a pyrophosphate transfer reaction that requires  $Mg^{2+}$  and proceeds by an ordered sequential mechanism.<sup>10-12</sup> Crystallographic studies of TPK were initiated to provide an improved understanding of

the enzyme and aspects of its structure and function at the molecular level.

## Results and Discussion

### Structure determination

Recombinant mouse TPK was crystallized by vapor diffusion (see Materials and Methods), and the crystal structure was determined at 2.4 Å (Figure 1(a)) from multiwavelength anomalous diffraction<sup>16</sup> data measured using selenomethionine substituted crystals (Table 1). Data collected



**Figure 1.** Electron density maps. (a) Experimental electron density calculated at 2.4 Å and contoured at 1.75  $\sigma$  is shown in stereo for residues to Gly225 to Ile241. (b) A 1.9 Å omit map was calculated following a 5000 K simulated annealing refinement using a model lacking thiamin. The thiazole and pyrimidine rings are numbered according to use in Results and Discussion. The  $2F_o - F_c$  (dark blue) and  $F_o - F_c$  (light blue) maps are contoured at 1.8 and 4.0  $\sigma$ , respectively. The Figure was created using O.<sup>33</sup>

Table 1. Mouse TPK data collection, structure solution and refinement statistics

	MAD Inflection	MAD Peak	MAD Remote	Thiamin
Energy (keV)	12.644	12.648	13.400	
Observations (total/unique)	273,244/49,984	282,176/49,817	284,364/50,110	330,681/52,413
Completeness (%)	100.0(100.0)	100.0(100.0)	100.0(100.0)	99.9(98.9)
$R_{\text{merge}}$	0.069(0.465)	0.063(0.337)	0.069(0.508)	0.073(0.406)
$I/\sigma I$	22.9(3.7)	25.5(5.3)	24.1(3.5)	18.6(2.2)
Resolution	30.0-2.4	30.0-2.4	30.0-2.4	30.0-1.9
Residues		496		496
Water molecules		255		413
Sulfate groups		0		14
Thiamin molecules		0		2
$R(R_{\text{free}})$		0.225(0.250)		0.214(0.239)
rmsd bonds (Å)		0.007		0.008
rmsd angles (deg.)		1.34		1.43
Ramachandran most favored (%)		89.1		90.8
Ramachandran allowed (%)		9.7		8.7

Energy in keV  $\approx 12.398/\lambda$  in Å. Values for MAD data are from SCALEPACK with Bijvoet pairs kept separate in scaling.  $R_{\text{merge}} = \Sigma |I - \langle I \rangle| / \Sigma I$ , where  $I$  is the integrated intensity of a given reflection. Numbers in parentheses represent the values obtained in the highest resolution bin (2.49-2.40 Å for MAD; 1.97-1.90 for thiamin). The refinement residual,  $R = \Sigma |F_{\text{obs}} - F_{\text{calc}}| / \Sigma F_{\text{obs}}$ , with the cross-validating  $R_{\text{free}}$  calculated from 1314 and 2606 test reflections excluded from refinements of the MAD and thiamin models, respectively.

from crystals soaked with thiamin were used for subsequent refinement of the enzyme/substrate complex at 1.9 Å (Figure 1(b)). The asymmetric unit of the TPK crystals contains two subunits. Electron density for the main chain is continuous for authentic residues in both subunits. However, disorder is indicated for the N-terminal His tags with nine residues missing from the electron density for one subunit and 19 missing from the other subunit.

### The TPK structure

The TPK subunit structure can be described as an N-terminal  $\alpha/\beta$ -domain and a C-terminal  $\beta$ -sandwich domain assembled about a central ten-stranded  $\beta$ -sheet (Figure 2). The first half of the central  $\beta$ -sheet is arranged as parallel  $\beta$ -strands sandwiched between  $\alpha$ -helices to form a three-layered  $\alpha/\beta/\alpha$  sandwich. The second half is formed from a jelly-roll of antiparallel strands which gives rise to the second sheet of the two-layered  $\beta$ -sandwich. The overall dimensions of the subunits are about 56 Å in length and 31 Å in width.

The TPK subunits associate as a dimer shaped like a diamond (Figure 2(c)) or a U (Figure 2(d)) when viewed parallel with or perpendicular to the non-crystallographic 2-fold axis, respectively. Residues have been numbered relative to the authentic start Met residue and end in A or B to distinguish between the two subunits. The dimeric association is mediated primarily by interactions between the  $\beta$ -sandwich domain of one subunit with both the  $\alpha/\beta$  and  $\beta$ -sandwich domains of the opposite subunit. These contacts primarily involve the loops between residues Pro207 and Pro211 and between Gly230 and Thr233 that protrude in towards the dimer interface from the opposing  $\beta$ -sandwich

domains. Limited direct contact is also made between the  $\alpha/\beta$ -domains in the vicinity of residues Leu128 to Leu139 and their 2-fold related counterparts. Consistent with previous studies indicating a dimeric association in solution,<sup>10,14</sup> the solvent-accessible surface area buried between the subunits is 2808 Å<sup>2</sup>. Surprisingly, 11 residues of the His tag from one subunit extend away from the TPK dimer to make a seemingly unlikely lattice contact that forms a short two-stranded  $\beta$ -sheet centered about a 2-fold crystallographic axis.

### Thiamin-binding site

The location of the TPK active site has been identified using data collected from crystals soaked with thiamin (Table 1; Figure 1(b)). The thiamin-binding site is located in the dimer interface distal to the non-crystallographic 2-fold axis in a cleft between the  $\alpha/\beta$ -domain of one subunit and the  $\beta$ -sandwich of the opposite subunit (Figure 2(c) and (d)). This binding site is formed by a  $\beta$ -bulge between Ser216B and Asn219B, and loops between Gly199B and Asn203B, and between Asp95A and Asp100A (Figure 3). The pyrimidine ring and methylene C atom stack parallel with the Trp202B indole ring (Figure 1(b)). The pyrimidine methyl group is in van der Waals contact with the side-chains of Leu204B and Leu214B in the hydrophobic core of the  $\beta$ -sandwich. Hydrogen bond potential exists for the side-chains of Ser216B and Asn219B and the pyrimidine ring N1' atom, while the Asp97A side-chain appears to accept a hydrogen bond from the amino group and participate in a water-mediated hydrogen bond to the N2' atom (Figure 3). The positive charge on the thiazole ring N3 atom may interact with partial negative charges on the main-chain O atoms of Thr21B and Gln96A located to either side. The thiazole methyl

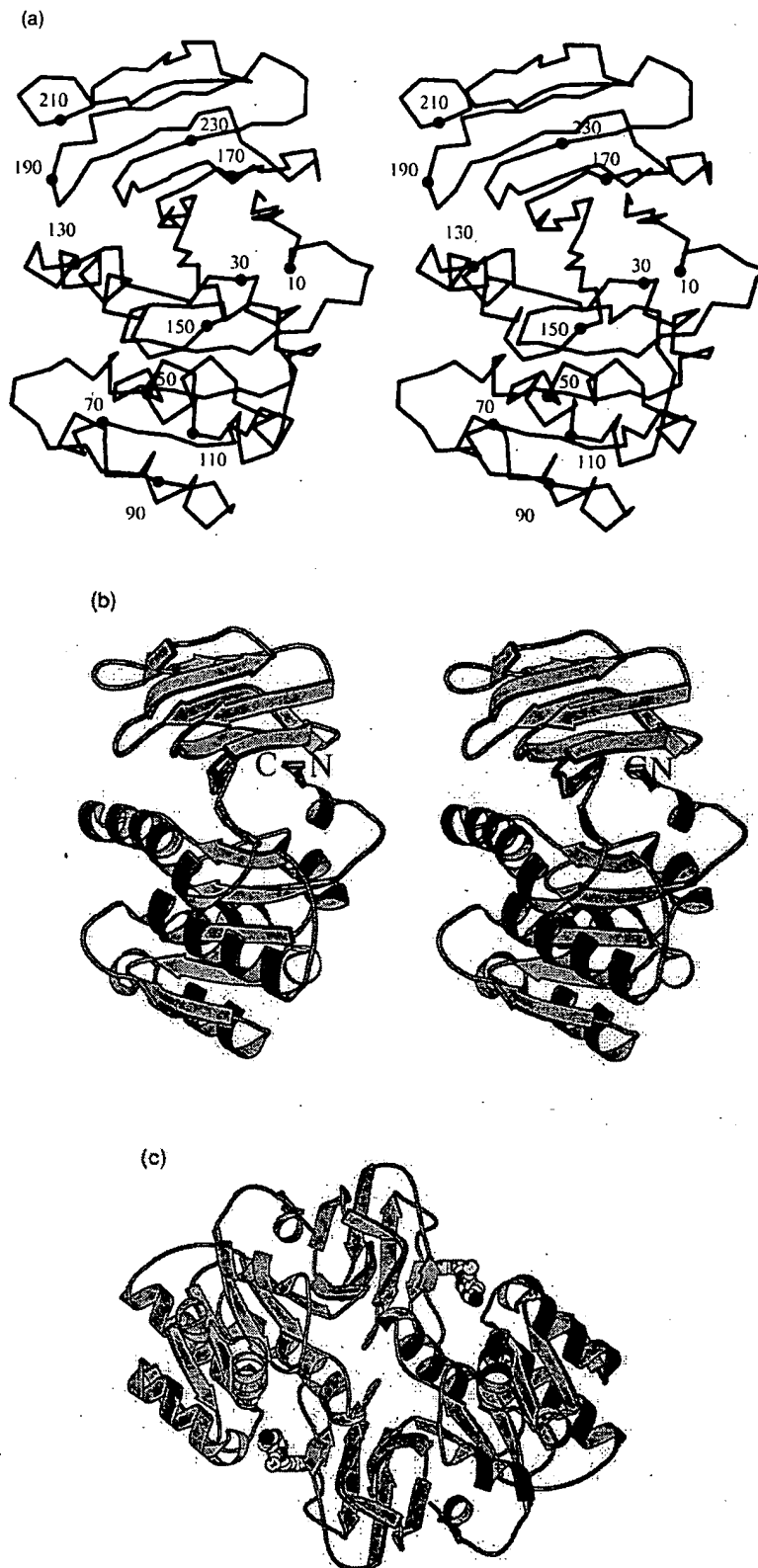
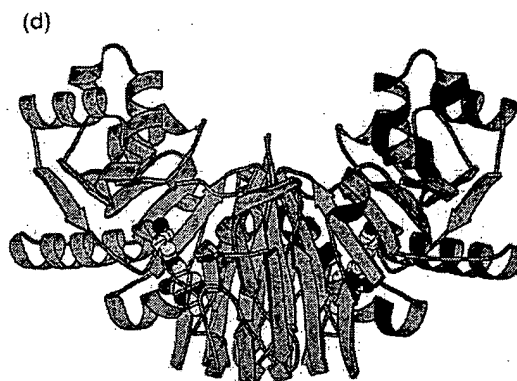


Figure 2(a)-(c) (legend opposite)



**Figure 2.** The TPK structure. (a) A  $C^\alpha$  trace is shown in stereo for the mouse TPK subunit. (b) A ribbon diagram illustrating the distribution of secondary structures within TPK is shown in stereo. (c) A ribbon diagram of the TPK dimer is viewed along the centrally located 2-fold non-crystallographic axis. (d) A ribbon diagram of the TPK dimer viewed perpendicular to the 2-fold non-crystallographic axis. The position of thiamin is represented in (c) and (d) as a ball-and-stick model with C, N, O and S atoms colored yellow, blue, red and green, respectively. This Figure and Figures 3 and 4 were created using MOLSCRIPT.<sup>36</sup>

group and C5 atom are also in van der Waals contact with the CH<sub>2</sub> atom of Trp202B and the C $\beta$  atom of Ser218B, respectively. The thiamin hydroxyethyl group points in the direction of the non-crystallographic 2-fold axis into a solvent-accessible space that occurs in the  $\alpha/\beta$ -domain between the parallel  $\beta$ -strands and the  $\alpha$ -helices beginning at Gly47A, Thr99A, and Arg131A.

Crystal structures of several TPP-dependent enzymes have been reported.<sup>17–20</sup> These include transketolase, pyruvate decarboxylase, pyruvate oxidase, benzoylformate decarboxylase and branched-chain  $\alpha$ -keto acid dehydrogenase. Crystal structures have also been reported for thiamin phosphate synthase<sup>21</sup> and hydroxyethylthiazole kinase,<sup>22</sup> which are involved in microbial thiamin biosynthesis. Some general features are shared in common between TPK and certain TPP-dependent enzymes. The pyrimidine ring tends to bind in a hydrophobic environment for most of these enzymes. Also, the intersubunit binding site and aromatic stacking of the pyrimidine ring in TPK are similar to interactions observed in transketolase and the  $\alpha$ -ketoacid dehydrogenase. However, these features are not universal among all TPP-dependent enzymes, and many differences can be identified. The TPP-dependent enzymes impose an energetically unfavorable "V" conformation between the pyrimidine and thiazole rings that positions the amino group to activate the C2 carbanion.<sup>23</sup> Thiamin phosphate synthase also imposes a V conformation, but places the thiazole methyl group in a *cis* position relative to the amino group.

In contrast, TPK has no apparent functional gain from unfavorable steric arrangements of the two ring systems and represents the first example of thiamin binding to an enzyme in the low-energy

"F" conformation. Thus, the highly conserved intercalation of a hydrophobic side-chain between the pyrimidine and thiazole rings in TPP-dependent enzymes and thiamin phosphate synthase is absent from TPK. Interactions of polar groups with both the pyrimidine and thiazole groups observed in the TPK/thiamin structure also vary significantly compared to other thiamin-binding enzymes. In particular, the conserved hydrogen bonding of a Glu side-chain with the pyrimidine N1' atom believed to stabilize an imino tautomer of N4' in TPP-dependent enzymes<sup>23</sup> is absent from the TPK structure.

### Comparison to yeast TPK

Mouse TPK has been compared to the structure of TPK from *Saccharomyces cerevisiae* to evaluate the conservation of structural features between enzymes sharing a 26% sequence identity over 288 residues. Details of the yeast TPK structure, containing thiamin and refined at 1.8 Å to an *R*-factor of 0.216 ( $R_{\text{free}} = 0.234$ ), will be reported elsewhere. The yeast TPK amino acid sequence is 75 residues longer than mouse TPK. This difference in size is due to 20 additional N-terminal and seven additional C-terminal residues, a 38-residue insertion and several smaller insertions of three residues or less. Overall, the TPK dimers from these species superimpose with a root-mean-square deviation (rmsd) of 1.6 Å for 406  $C^\alpha$  atoms in 19 secondary structures (Figure 4(a)). The N-terminal extension in yeast TPK wraps around the  $\beta$ -sandwich domain to add an additional antiparallel  $\beta$ -strand in close proximity to the thiamin binding site. The 38 residue insertion occurs at the end of an  $\alpha$ -helix to replace a short loop between Lys115 and Val121 in mouse TPK with a structure containing two  $\alpha$ -

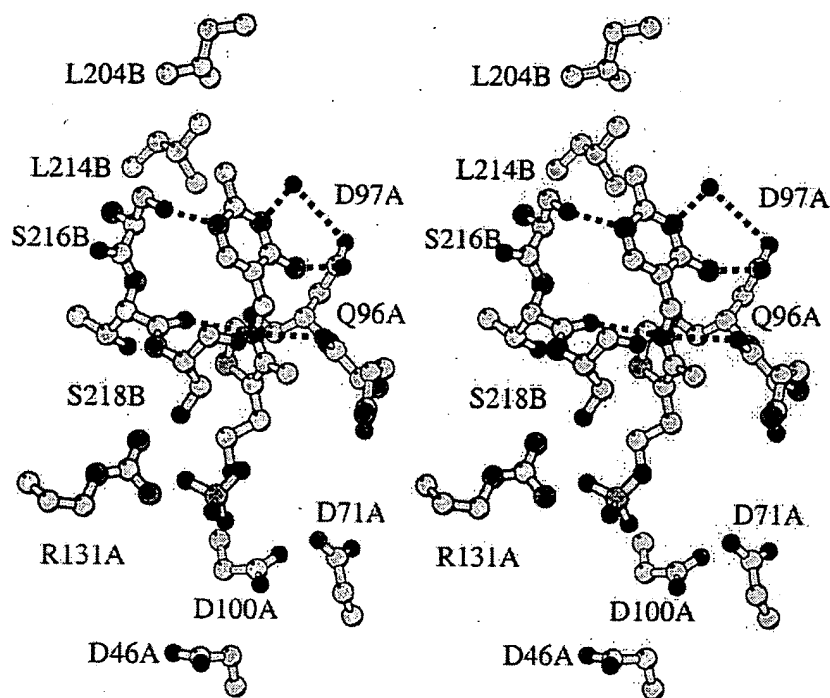


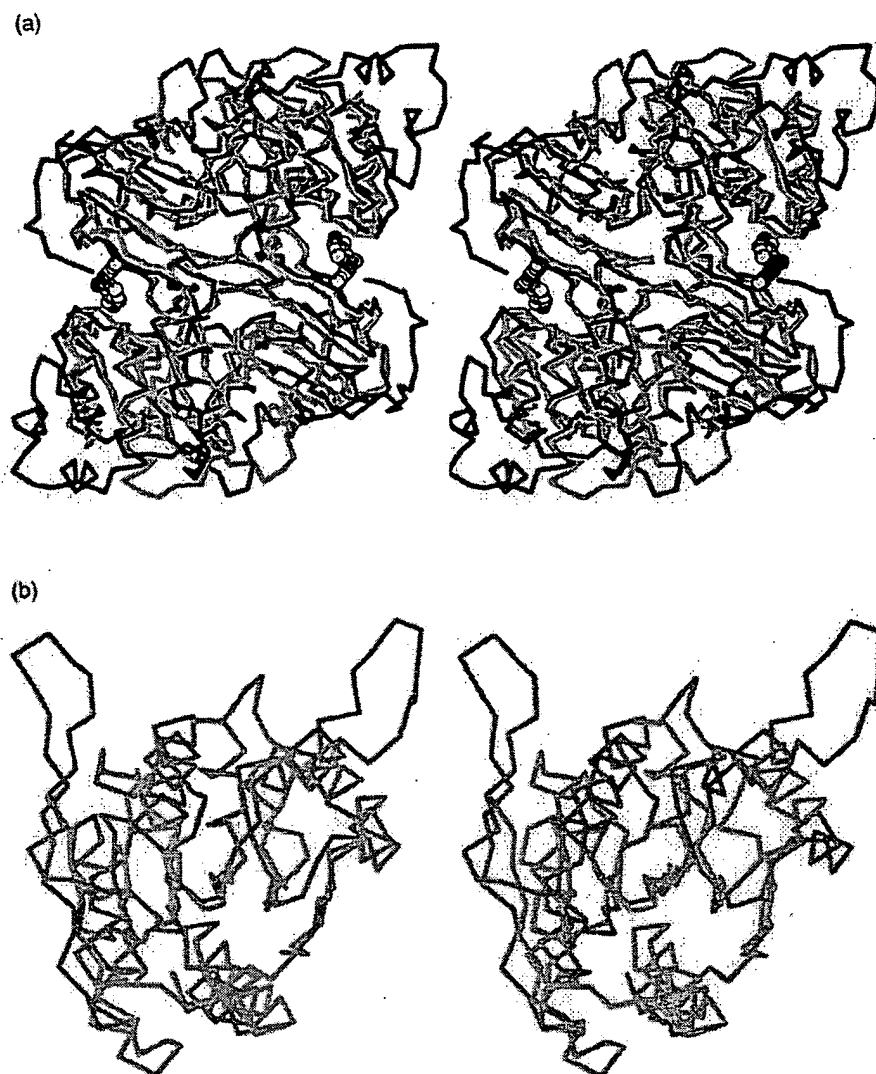
Figure 3. The TPK active site. A stereodigram of the TPK active site is shown with potential electrostatic interactions described in Results and Discussion indicated by broken lines. The Trp202 side-chain, located in the right foreground of the Figure, has been omitted to maintain an unobstructed view. Atom coloring has been described in the legend to Figure 2.

helices. This insertion packs against the  $\alpha/\beta$ -domain near the helix between Thr99 and Lys115 and the edge of the parallel  $\beta$ -sheet in close proximity to the loop contacting the thiazole ring between Asp95 and Asp100. Many of the residues in the active site and contacts between thiamin and TPK are conserved, including the pyrimidine ring stacking against a Trp side-chain, the hydrogen bonding of a Ser side-chain to the N1' atom, and the positioning of the thiazole N3 atom between main-chain oxygen atoms (Figure 3). However, the environment surrounding the pyrimidine methyl group in mouse TPK is formed by a Cys, a Trp and a Val residue in yeast TPK, and the Asp interacting with the N3' and N4' atoms in mouse TPK is replaced by Tyr.

#### Other similar structures

While structures similar to both TPK domains are common in the Protein Data Bank (PDB), it is worth noting that these motifs are also present in other proteins of related function. Structures resembling the TPK  $\alpha/\beta$ -domain are present in another pyrophosphokinase and several small molecule kinases. Repeats of a five-stranded parallel  $\beta$ -

sheet flanked by helices on either side are found in the structure of phosphoribosylpyrophosphate (PRPP) synthetase.<sup>24</sup> This enzyme also catalyzes a pyrophosphate transfer reaction and superimposes on mouse TPK with an rmsd of 2.0 Å over 58 C $\alpha$  atoms present in five secondary structures (Figure 4(b)). Comparable levels of structural similarity are also found for other small molecule kinases, including adenosine kinase, ribokinase and hydroxyethylthiazole kinase (ThiK).<sup>22</sup> Interestingly, ThiK is required for microbial biosynthesis of thiamin *via* a salvage pathway, and the active sites of TPK and ThiK occur in the same general vicinity following superposition. However, the substrate binding sites do not superimpose between the two structures. The hydroxyethyl thiazole ring in ThiK is located about 9 Å away from the thiamin thiazole ring in TPK, and the superimposed ATP bound to ThiK clashes with the TPK loop between Asp95 and Asp100. Besides PRPP, the only other known pyrophosphokinase structure of hydroxymethyldihydropterin pyrophosphokinase (HPPK)<sup>25</sup> shows only superficial similarity to TPK, with a three-layered  $\alpha/\beta/\alpha$  structure containing a central antiparallel rather than a parallel  $\beta$ -sheet. Finally, elements of the TPK  $\alpha/\beta$ -domain



**Figure 4.** TPK superpositions. (a) The mouse TPK dimer (blue) is superimposed with the yeast homolog (red). The view is onto the "back side" of the active site with the structures rotated approximately  $180^\circ$  in the plane of the paper relative to the view shown in Figure 2(c). Thiamin molecules are colored as described in the legend to Figure 2. (b) The TPK  $\alpha/\beta$ -domain (blue) is superimposed on a similar structure present in phosphoribosylpyrophosphate synthetase (violet).

also resemble the ubiquitous Rossmann fold present in many nucleotide-binding proteins.<sup>26</sup>

Numerous structural homologs of the TPK  $\beta$ -sandwich domain are also represented in the PDB. The TPK jelly-roll motif is most similar to a number of viral structural proteins and conalbumins. Similarities between TPK and an epimerase, RmlC,<sup>27</sup> from Gram-negative bacteria may be of greater functional significance. The superposition of RmlC and TPK gives an rmsd of 1.7 Å for 39 C $\alpha$  atoms in six secondary structures. The pyrimidine ring of the nucleotide substrate of RmlC binds to the exposed edge of a  $\beta$ -sandwich, stacking against a Tyr

side-chain in a manner similar to the pyrimidine binding to TPK. However, the comparison is again general in nature, and the RmlC binding site actually occurs on the opposite side of the  $\beta$ -sandwich relative to the superimposed TPK structure.

### C nclusions

The TPK structure provides a molecular framework for understanding the general mechanism and regulation of an essential enzyme and pyrophosphate transfer reactions. This type of biochemical reaction is essential for many aspects of

living systems. In addition to thiamin metabolism, pyrophosphate transfer reactions are also essential for the biosynthesis of amino acids, folates, purine and pyrimidine bases, and the nucleotides and coenzymes derived from these compounds. Recent reports of pyrophosphokinase structures provide an opportunity to further the understanding of this group of enzymes, which lags significantly behind traditional kinases that transfer orthophosphate.

While additional studies of TPK are required to determine details of ATP and  $Mg^{2+}$  binding, regulation by TPP and inhibition by ethanol, the TPK/thiamin structure provides some insight into aspects of TPK function. A likely ATP-binding site is formed in the vicinity of thiamin and the Arg131 side-chain. The high degree of solvent accessibility of this site is consistent with the relatively high  $K_m$  value reported for ATP in kinetic studies of TPK.<sup>10-12</sup> An interaction of the  $\beta$  and/or  $\gamma$  phosphate groups with the guanidinium group would place the  $\beta$ -phosphorus atom in a position suitable for in-line nucleophilic attack by the thiamin hydroxyl oxygen atom. The structure and sequence surrounding Arg131 is consistent with this scenario. A sulfate ion is bound to the Arg131 side-chain located within 2.5 Å of the thiamin hydroxyl (Figure 3). A nearby cluster of Asp residues (46, 71, 73 and 100) is likely to function along with phosphate groups and potentially the thiamin hydroxyl group in the coordination of the  $Mg^{2+}$  cofactor required for catalytic activity. Asp side-chains also coordinate  $Mg^{2+}$  in many TPP-dependent enzymes. Arg131 is at the N terminus of an  $\alpha$ -helix preceded by a loop containing a Gly-rich sequence (126-GGLGGR-131) reminiscent of the Rossmann fold<sup>26</sup> GXGXXG motif that functions with the helix dipole to bind nucleotide phosphate groups. A Gly-rich motif (14-GACKGT-19) also interacts with nucleotide bound to adenosine kinase.<sup>28</sup> However, barring a large conformational change, the 13.5 Å distance between thiamin and the nearest Gly residue in this loop makes it unlikely for the Gly residues to directly contact phosphate groups in TPK. This situation is similar to that of aldehyde dehydrogenase,<sup>29</sup> which revealed novel placement and use of the Rossmann fold in nucleotide binding. It should also be noted that the Gly-rich motif, Arg131 and the Asp cluster are strictly conserved between TPKs from different species. However, the lack of sequence similarities and conservation of specific ligand interactions between TPK and other thiamin and nucleotide binding proteins suggests that rather than having an evolutionary relationship, common structural motifs have simply been adapted in different ways to a common use.

Finally, the results of these crystallographic studies suggest that TPK should be considered as a target for the design of antimicrobial agents. TPK is essential for basic carbohydrate utilization; *S. cerevisiae* and *Schizosaccharomyces pombe* growth is dependent on TPK,<sup>2,7,8</sup> and other pyrophosphokinases<sup>24,25</sup> are being targeted for similar purposes. Mouse TPK residues Asp97,

Asp207 and Leu214 are substituted in yeast by Tyr, Trp and Arg, respectively. These non-conservative substitutions could be exploited in developing species-specific thiamin analogs. Furthermore, the

additional N-terminal  $\beta$ -strand present in yeast TPK places a sulfhydryl group in van der Waals contact with thiamin. While additional full-length TPK sequences will need to be determined to fully assess TPK as a target, the mouse enzyme shares 91% identity over 137 residues of a partial human sequence, but only 38% identity with the next closest sequence available. Therefore, it seems likely that the level of variability between mouse and yeast TPK will be similar to the levels between humans and pathogens.

## Materials and Methods

### Expression and crystallization

Mouse TPK<sup>15</sup> has been cloned and expressed in *Escherichia coli* as a polyhistidine fusion protein using the vector pET28a. The fusion protein contains a 20 residue His tag for a total of 263 residues. TPK was purified using nickel-charged metal chelate chromatography (Pharmacia) and elution with imidazole, and is stored frozen at  $-80^\circ\text{C}$  in 150 mM NaCl. Crystals were grown at room temperature in the space group P3<sub>1</sub>21 having unit cell dimensions of 90.1 Å × 90.1 Å × 140.4 Å using the hanging drop vapor diffusion method where equal volumes of an 11 mg/ml TPK solution were mixed with precipitant solutions containing 1.8-2.1 M ammonium sulfate, 0.1 M Hepes and 2-4% PEG400, pH 7.1.

### Structure determination and crystal soaking

The TPK crystal structure has been determined using the multiwavelength anomalous diffraction (MAD) method.<sup>16</sup> Selenomethionine-substituted TPK was produced by over-expressing the recombinant enzyme in a defined medium containing selenomethionine and using the methionine auxotrophic strain, DL41(DE3). MAD data (Table 1) were collected at the Advanced Photon Source synchrotron beam line 19ID. The data were integrated using HKL2000 and merged and scaled using SCALEPACK.<sup>30</sup> The partial structure of seven selenium atoms was determined using the program SOLVE.<sup>31</sup> The Se parameters were further refined and crystallographic phases were calculated to 2.4 Å using MLPHARE.<sup>32</sup> Further improvements of the experimental electron density map were achieved by density modification and 2-fold non-crystallographic symmetry averaging implemented in DM.<sup>32</sup> Data from TPK crystals soaked in precipitant solution containing 10 mM thiamin for 16 hours were collected using a Rigaku RU200 rotating anode X-ray generator and an R-axis IIC detector. These data were integrated using HKL v1.96.0 and merged and scaled using SCALEPACK.<sup>30</sup>

### Model building and refinement

The initial SeMet-substituted TPK model containing 494 amino acid residues was built using the program O.<sup>33</sup> This model was refined with cycles of water building using REFMAC<sup>32</sup> and ARPP<sup>32</sup> with non-crystallographic symmetry restraints and a bulk solvent

correction to yield a model having 66 water molecules and an  $R$ -factor of 0.26 ( $R_{\text{free}} = 0.29$ ). Manual rebuilding of the model, placement of additional water molecules and refinement using CNS<sup>34</sup> v0.9 with non-crystallographic symmetry restraints and a bulk solvent correction yielded an  $R$ -factor of 0.225 ( $R_{\text{free}} = 0.250$ ) at 2.4 Å (Table 1). This model was used in three additional cycles of model building and refinement against 1.9 Å data collected from the thiamin-soaked crystal. Structural superpositions were performed using TOP.<sup>35</sup>

### Coordinates

Coordinates and structure factors have been deposited in the RCSB Protein Data Bank under the accession code 1IG3.

### Acknowledgments

The authors thank Norma Duke and Greg Titus for assistance in data collection, Xinhua Ji for sharing coordinates of HPPK prior to release, and Tom Hurley and Nic Steussy for useful discussions. R.A.H. was supported by NIH grant DK19259. D.E.T. was supported by NIH grant DK54738 and a grant from the Indiana Alcohol Research Center. L.J.B. was supported by a post-doctoral fellowship from the Midwest Affiliate of the American Heart Association.

### References

- Gubler, C. J. (1990). Thiamin. In *Handbook of Vitamins* (Machlin, L., ed.), Marcel Dekker, Inc. New York, Basel.
- Hohmann, S. & Meacock, P. A. (1998). Thiamin metabolism and thiamin diphosphate-dependent enzymes in the yeast *Saccharomyces cerevisiae*: genetic regulation. *Biochim. Biophys. Acta*, 1385, 201-219.
- Jordan, F. (1999). Interplay of organic and biological chemistry in understanding coenzyme mechanisms: example of thiamin diphosphate-dependent decarboxylations of 2-oxo acids. *FEBS Letters*, 457, 298-301.
- Tallaksen, C. M. E., Bell, H. & Bohmer, T. (1993). Thiamin and thiamin phosphate ester deficiency assessed by HPLC in four clinical cases of Wernicke encephalopathy. *Alcoholism: Clin. Exp. Res.* 17, 712-716.
- Rindi, G., Imarisio, L. & Patrini, C. (1986). Effects of acute and chronic ethanol administration of regional thiamin pyrophosphokinase activity of the rat brain. *Biochem. Pharm.* 35, 3903-3908.
- Laforenza, U., Patrini, C., Gastaldi, G. & Rindi, G. (1990). Effects of acute and chronic ethanol administration on thiamine metabolizing enzymes in some brain areas and in other organs of the rat. *Alcohol Alcoholism*, 25, 591-603.
- Nosaka, K., Kaneko, Y., Nishimura, H. & Iwashima, A. (1993). Isolation and characterization of a thiamin pyrophosphokinase gene, THI80, from *Saccharomyces cerevisiae*. *J. Biol. Chem.* 268, 17440-17447.
- Fankhauser, H., Zurlinden, A., Schweingruber, A.-M., Edenharter, E. & Schweingruber, M. E. (1995). *Schizosaccharomyces pombe* thiamine pyrophosphokinase is encoded by gene *trr3* and is a regulator of thiamine metabolism, phosphate metabolism, and growth. *J. Biol. Chem.* 270, 28457-28462.
- Lau, K. S., Fatania, H. R. & Randle, P. J. (1982). Regulation of the branched chain 2-oxoacid dehydrogenase kinase reaction. *FEBS Letters*, 144, 57-62.
- Voskoboyev, A. I. & Ostrovsky, Y. M. (1982). Thiamin pyrophosphokinase: structure, properties and role in thiamin metabolism. *Ann. Acad. Sci.* 378, 161-176.
- Gubler, C. J., Fleming, G. & Kuby, S. A. (1996). Thiamin pyrophosphokinase. Purification, properties, and function. In *Biochemistry and Physiology of Thiamin Diphosphate Enzymes* (Bisswanger, H. & Schellenberger, A., eds), pp. 557-5569, A.U.C. Intemann, Wissenschaftlicher Verlag, Prien.
- Mitsuda, H., Takii, Y., Iwami, K. & Yasumoto, K. (1975). Purification and properties of thiamine pyrophosphokinase from parsley leaf. *J. Nutr. Sci. Vitaminol.* 21, 103-115.
- Matsuda, T. & Cooper, J. R. (1981). Thiamine as an integral component of brain synaptosomal membranes. *Proc. Natl Acad. Sci. USA*, 78, 5885-5889.
- Sanemori, H. & Kawasaki, T. (1980). Purification and properties of thiamine pyrophosphokinase. *J. Biochem.* 88, 223-230.
- Nosaka, K., Onozuka, M., Nishino, H., Nishimura, H., Kawasaki, Y. & Ueyama, H. (1999). Molecular cloning and expression of a mouse thiamin pyrophosphokinase cDNA. *J. Biol. Chem.* 274, 34129-34133.
- Yang, W., Hendrickson, W. A., Crouch, R. J. & Satow, Y. (1990). Structure of ribonuclease H phased at 2 Å resolution by MAD analysis of the selenomethionyl protein. *Science*, 249, 1398-1405.
- Muller, Y. A., Lindqvist, Y., Furey, W., Schulz, G. E., Jordan, F. & Schneider, G. (1993). A thiamin diphosphate binding fold revealed by comparison of the crystal structures of transketolase, pyruvate oxidase and pyruvate decarboxylase. *Structure*, 1, 95-103.
- Hasson, M. S., Muscate, A., McLeish, M. J., Polovnikova, L. S., Gerlt, J. A., Petsko, G. A. & Ringe, D. (1998). The crystal structure of benzoylformate decarboxylase at 1.6 Å resolution: diversity of catalytic residues in thiamin diphosphate-dependent enzymes. *Biochemistry*, 37, 9918-9930.
- Aevarsson, A., Seger, K., Turley, S., Sokatch, J. R. & Hol, W. G. (1999). Crystal structure of 2-oxoisovalerate dehydrogenase and the architecture of 2-oxo acid dehydrogenase multienzyme complexes. *Nature Struct. Biol.* 6, 785-792.
- Lindqvist, Y., Schneider, G., Ermler, U. & Sundström, M. (1992). Three-dimensional structure of transketolase, a thiamine diphosphate dependent enzyme, at 2.5 Å resolution. *EMBO J.* 11, 2373-2379.
- Chiu, J. J., Reddick, J. J., Begley, T. P. & Ealick, S. E. (1998). Crystal structure of thiamin phosphate synthase from *Bacillus subtilis* at 1.25 Å resolution. *Biochemistry*, 38, 6460-6470.
- Campobasso, N., Mathews, I. I., Begley, T. P. & Ealick, S. E. (2000). Crystal structure of 4-methyl-5-beta-hydroxyethylthiazole kinase from *Bacillus subtilis* at 1.5 Å resolution. *Biochemistry*, 39, 7868-7877.
- Kern, D., Kern, G., Neef, H., Tittmann, K., Killenberg-Jabs, M., Wikner, C., Schneider, G. & Hübner, G. (1997). How thiamine diphosphate is activated in enzymes. *Science*, 275, 67-70.
- Eriksen, T. A., Kadziola, A., Bentsen, A.-K., Harlow, K. W. & Larsen, S. (2000). Structural basis for the function of *Bacillus subtilis* phosphoribosyl-pyrophosphate synthetase. *Nature Struct. Biol.* 7, 303-308.

25. Xiao, B., Shi, G., Chen, X., Yan, H. & Ji, X. (1999). Crystal structure of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase, a potential target for the development of novel antimicrobial agents. *Structure*, **7**, 489-496.
26. Rossmann, M. G., Moras, D. & Olsen, K. W. (1974). Chemical and biological evolution of a nucleotide binding protein. *Nature*, **250**, 194-199.
27. Giraud, M.-F., Leonard, G. A., Field, R. A., Berlind, C. & Naismith, J. H. (2000). RmlC, the third enzyme of dTDP-L-rhamnose pathway, is a new class of epimerase. *Nature Struct. Biol.* **7**, 398-402.
28. Abele, U. & Schulz, G. E. (1995). High-resolution structures of adenylate kinase from yeast ligated with inhibitor Ap5A, showing the pathway of phosphoryl transfer. *Protein Sci.* **4**, 1262-1271.
29. Liu, Z. J., Sun, Y. J., Rose, J., Chung, Y. J., Hsiao, C. D., Chang, W. R., Kuo, I., Perozich, J., Lindahl, R., Hempel, J. & Wang, B. C. (1997). The first structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossmann fold. *Nature Struct. Biol.* **4**, 317-326.
30. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation data reduction mode. *Methods Enzymol.* **276**, 307-326.
31. Terwilliger, T. C. & Berendzen, J. (1999). Automated structure solution for MIR and MAD. *Acta Crystallog. sect. D*, **55**, 849-861.
32. Collaborative Computing Project No. 4 (1994). The CCP4 suite of programs. *Acta Crystallog. sect. D*, **50**, 760-763.
33. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110-119.
34. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905-921.
35. Lu, G. (2000). TOP: a new method for structure comparisons and similarity searches. *J. Appl. Crystallog.* **33**, 176-183.
36. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.

Edited by M. F. Summers

(Received 27 February 2001; received in revised form 4 April 2001; accepted 4 April 2001)

# Chemical and biological evolution of a nucleotide-binding protein

Michael G. Rossmann, Dino Moras & Kenneth W. Olsen

Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907

*Three-dimensional alignment of the common nucleotide binding structure in dehydrogenases, kinases and flavodoxins permits the recognition of homologous amino acids when sequence comparisons alone would fail. Minimum base changes per codon can then be used to measure evolutionary distances which suggest that this structure was present during precellular evolution.*

A COMMON structural domain<sup>1</sup> whose function is to bind nicotinamide adenine dinucleotide (NAD) has been found in lactate dehydrogenase (LDH)<sup>2,3</sup>, in soluble malate dehydrogenase (sMDH)<sup>4</sup>, in liver alcohol dehydrogenase (LADH)<sup>5</sup> and in glyceraldehyde-3-phosphate dehydrogenase (GAPDH)<sup>6</sup>. Rao and Rossmann<sup>7</sup> showed that the same structure was utilised to bind flavin mononucleotide (FMN) in flavodoxin<sup>8-10</sup>. They also showed that this structure consists of two smaller units the function of each being to bind a mononucleotide. Buehner *et al.*<sup>6</sup> proposed an evolutionary tree which traces the incorporation and evolution of the mononucleotide-binding structure in various dehydrogenases and in flavodoxin. Further data on the three-dimensional structures and amino acid sequences of some dehydrogenases and flavodoxins are now available which permits the evaluation of the proposed evolutionary tree more precisely. We show here that the position and sequence of the nodes in this tree are consistent with the molecular data, and that a rough time scale for these events can be proposed.

Independent support for the evolutionary tree comes from a study of redox potentials and the evolution of biological electron transport<sup>11,12</sup>, suggesting a common origin for NAD and flavin-binding proteins. Furthermore, structures similar to the mononucleotide or dinucleotide-binding protein fragment have been found in phosphoglycerate kinase<sup>13</sup> (Bryant *et al.*<sup>14</sup> suggest an alternative solution) and tentatively in adenyl kinase<sup>15</sup>. Both these enzymes need to bind AMP, ADP or ATP. Similarly a flavodoxin-like structure has been recognised tentatively in rhodanese<sup>16</sup> whose biological function is not fully established but is known to bind flavin FAD FMN and NAD.

## Structural alignments

The nucleotide-binding protein discussed above consists of a parallel sheet formed by three extended polypeptide strands. The first two are connected by an  $\alpha$ -helix, and the last two by

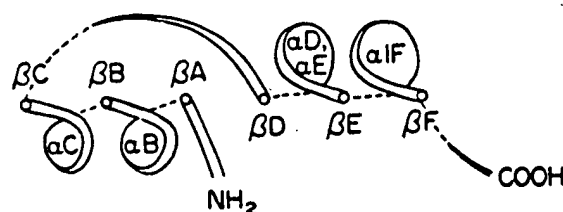


Fig. 1 Diagrammatic representation of a dinucleotide-binding protein. The amino termini of the strands in the  $\beta$ -pleated sheet are closest to the viewer. The dinucleotide binds to the carboxy termini of the strands in the sheet.

either an  $\alpha$ -helix or another, less well defined, structure. In dehydrogenases the secondary structural features of the dinucleotide-binding fragment have been termed  $\beta$ A,  $\beta$ B,  $\beta$ C,  $\beta$ D,  $\beta$ E and  $\beta$ F within the  $\beta$ -pleated sheet, while the helices are labelled  $\alpha$ B,  $\alpha$ C, . . . (Fig. 1). This nomenclature was originally devised for the LDH structure<sup>3</sup>. Helices  $\alpha$ B and  $\alpha$ C connect strands  $\beta$ A with  $\beta$ B and  $\beta$ B with  $\beta$ C, respectively. They are on the same side of the sheet. Along the polypeptide chain, the sequence  $\beta$ A,  $\alpha$ B,  $\beta$ B,  $\alpha$ C,  $\beta$ C forms the first (adenine) nucleotide-binding fragment, while the second (nicotinamide) nucleotide-binding fragment is related by an approximate two-fold axis. The fold of each fragment has the same unique band.

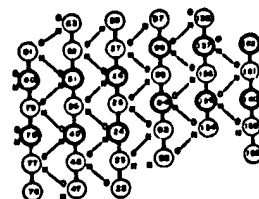
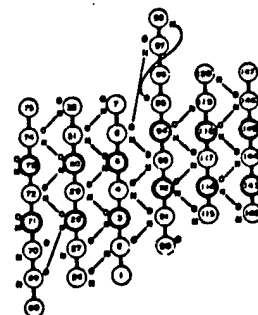


Fig. 2 Hydrogen bonding diagram of  $\beta$ -pleated sheet region in (top) dogfish apo-LDH and (bottom) lobster holo-GAPDH. Amino acids with thick outlines have residues facing into hydrophobic pockets between the  $\beta$ -pleated sheet and  $\alpha$  helices. The hydrophobic character of these residues is strongly conserved (Table 1).



There is a greater conservation of the structure of the sheet region than of the helices. For example, in LADH and GAPDH the helix  $\alpha$ D is missing, while in LDH and s-MDH it forms part of a flexible loop which undergoes a large conformational change during catalysis. The greater conservation of the parallel pleated sheet region can be seen readily in Fig. 2 where the hydrogen bonding within the sheet has been depicted for dogfish apo-LDH and lobster holo-GAPDH. When these bonds are equivalently aligned, with due regard for the direction of polarity of the peptide chain, then the orientation of the  $C_{\alpha}$  atom in each amino acid side chain must also be aligned. Hence superposition of the hydrogen bonding provides a sensitive method of determining homologous amino acids in equivalent  $\beta$ -sheet regions of different nucleotide-binding fragments. The atomic coordinates of the equivalenced  $C_{\alpha}$  atoms were taken to set up an approximate rotation matrix to superimpose the two molecules being compared. This was then refined to minimise the sum of the square of the distance between equivalenced atoms<sup>7</sup>. With this alignment other residues, not in the sheet, but nevertheless following the same

Table 1 Nucleotide-binding protein fragment comparison of amino acids

Reference to sequence data	BA	αB	βB	βC	
Dogfish LDH	38 2 N K I T V V G C B A V G 3 2 N K I T V V G C B A V G 3	3 M A D A I S V L M K D L A 4 4 M A D A I S V L M K D L A 4	4 D E V A L V D 3 3 D E V A L V D 3	7 A K I V S G K 8 8 A K I V S G K 8	
Pig GAPDH	39 1 K V G V D D G F F G R I G 2 2 K V G V D D G F F G R I G 2	1 L V T R A A A L S C G P B 9 9 L V T R A A A L S C G P B 9	1 V A I N D D P P P F I A L 3 3 V A I N D D P P P F I A L 3	6 K A I T I F F Q N E E 7 7 K A I T I F F Q N E E 7	
Lobster GAPDH	40 S K I G I D D G F F G R I G 2 2 S K I G I D D G F F G R I G 2	2 R L V M R I A L S R P B 2 2 R L V M R I A L S R P B 2	2 V V A S B B 2 2 V V A S B B 2	2 K K I A T T Y Q 2 2 K K I A T T Y Q 2	
Yeast GAPDH	41 1 V R V A I D D G F F G R I G 2 2 V R V A I D D G F F G R I G 2	1 R L V M R I A L S R P B 2 2 R L V M R I A L S R P B 2	1 V V A S B B 2 2 V V A S B B 2	1 K K I A T T Y Q 2 2 K K I A T T Y Q 2	
Horse LADH	42, 43 3 T C A V F G L L G G V G 2 2 T C A V F G L L G G V G 2	3 L S V I M G C K A A G A 2 2 L S V I M G C K A A G A 2	3 A R I I G V D 2 2 A R I I G V D 2	3 G A T E C V N P 3 3 G A T E C V N P 3	
Bovine GluDH	19 K T F A V Q G F F G N V G 1 1 K T F A V Q G F F G N V G 1	1 H S M R Y L H R F G A 2 2 H S M R Y L H R F G A 2	1 K C V A V G E 3 3 K C V A V G E 3	1 D G S I 1 1 D G S I 1	
Clostridium MP	10, 44, 45 1 M K I V Y W S G T G 1 1 M K I V Y W S G T G 1	1 E L I A K G I I E S G 1 1 E L I A K G I I E S G 1	1 K D V N T I N V S D V N 1 1 K D V N T I N V S D V N 1	1 P S V I A V G A 1 1 P S V I A V G A 1	
Flavodoxin					
Subtilisin	46 D V I N M S L G P S G 1 1 D V I N M S L G P S G 1	1 S A A L K A A V D K A G 7 7 S A A L K A A V D K A G 7	1 V V V V A A A 4 4 V V V V A A A 4	1 G N E G S P 2 2 G N E G S P 2	
Dogfish LDH	38 8 A G S K L V V I T A G A R Q Q 9 9 A G S K L V V I T A G A R Q Q 9	1 F K F I I P N I V K H S P D 1 1 F K F I I P N I V K H S P D 1	1 C I L E L H P 1 1 C I L E L H P 1	1 H R I I G B G 1 1 H R I I G B G 1	
Pig GAPDH	39 8 A G T A E V I A I V V V E S 9 9 A G T A E V I A I V V V E S 9	1 M E K A S A H L K K G G A 2 2 M E K A S A H L K K G G A 2	1 K R V I I S A A 4 4 K R V I I S A A 4	1 K I V S N A S S C C 8 8 K I V S N A S S C C 8	
Lobster GAPDH	40 A G A E V I A I V V V E S 2 2 A G A E V I A I V V V E S 2	1 T T E L D T A Q K H L K 2 2 T T E L D T A Q K H L K 2	1 K K V V I S A A 4 4 K K V V I S A A 4	1 L K I V S N A S S C C 3 3 L K I V S N A S S C C 3	
Yeast GAPDH	41 2 G D S 2 6 6 0 G V D F S F E V I G R 3 3 G V D F S F E V I G R 3	1 M V T A L S C C Q E A Y 7 7 M V T A L S C C Q E A Y 7	1 G V S V I V G 8 8 G V S V I V G 8	1 R T W K G A I 1 1 R T W K G A I 1	
Horse LADH	42, 43 4 G V D F S F E V I G R 3 3 G V D F S F E V I G R 3	1 M V T A L S C C Q E A Y 7 7 M V T A L S C C Q E A Y 7	1 G V S V I V G 8 8 G V S V I V G 8	1 R T W K G A I 1 1 R T W K G A I 1	
Clostridium MP	10, 44, 45 4 L N E D I L I L G C S A M G D 6 6 L N E D I L I L G C S A M G D 6	1 H E P F I E E I S T K I S G 7 7 H E P F I E E I S T K I S G 7	1 K K V A L F G 8 8 K K V A L F G 8	1 C V V V G T P L I 5 5 C V V V G T P L I 5	
Flavodoxin					

The AMP binding fragment above is aligned below with the NMN binding fragment in dehydrogenases and the FMN binding fragment in flavodoxin. LDH Cys 165 and GAPDH Cys 149 form the essential thiol groups of these two enzymes.

Table 2 Comparison of the NAD-binding domain among various dehydrogenases

		1	2	3	4
Dogfish LDH	1		1.12	1.16	1.13
Lobster GAPDH	2	75		1.13	1.00
Horse LADH	3	74	71		1.13
Bovine GluDH	4				

The top right of the matrix shows the minimum base changes per codon for the alignments in Table 1. The bottom left shows the number of those equivalent amino acids whose C $\alpha$  atoms approach each other to within 3.8 Å.

folding sequence could be equivalenced and incorporated into the refinement procedure.

The amino acid alignments shown in Table 1 are thus based on three-dimensional structure, while the character of the amino acids has been determined chemically. Since Rao and Rossman<sup>7</sup> have pointed out similarity of structure between a part of subtilisin<sup>17,18</sup> and a mononucleotide-binding fragment, with the aromatic specificity pocket of subtilisin corresponding to the adenine-binding pocket of the dehydrogenases, this comparison has been included in Table 1. Results for individual comparisons among the dehydrogenases are shown as a matrix in Table 2. The resultant alignment of the bound nucleotides is best seen in a series of stereographic diagrams. The C $\alpha$

backbone of the NAD-binding fragment of LDH has been used as a standard of comparison. In Fig. 3, LDH is compared with the dehydrogenases GAPDH and LADH, in Fig. 4 comparisons are made with phosphoglycerate kinase<sup>13</sup> and flavodoxin and Fig. 5 shows a comparison of the AMP-binding fragment with the MNN-binding fragment in LDH.

### Sequence homologies with GluDH

The alignment of the glutamate dehydrogenase (GluDH) sequences<sup>19</sup> is not based on structure but on amino acid sequence homologies alone. Different methods of testing sequences for homologies in distantly related proteins have been reviewed by Haber and Koshland<sup>20</sup>, Barker and Dayhoff<sup>21</sup> and others. The procedure described by Jukes and Cantor<sup>22</sup>, which depends on counting minimum base changes per codon, has been used here. By far the best alignment was found in comparing residues 1 to 38 in GAPDH with 245 to 283 in bovine GluDH corresponding to  $\beta$ A,  $\alpha$ B and  $\beta$ B (Table 1). This corresponds to the most conserved part of the NAD-binding structure and furthermore the character of important amino acids was maintained.

The glycine in position 28 of LDH is strictly conserved as any larger residue would cause steric hinderance to the binding of the ribose ring. The cause of conservation of the glycine in position 33 of LDH is probably related to its position on helix  $\alpha$ B immediately opposite the  $\beta$ -pleated sheet. The conservation of LDH aspartate 53 (changed to glutamate in GluDH) at the end of  $\beta$ B must be related to its function of binding the O2' atom in the adenine ribose<sup>3</sup>. The alternate hydrophobic residues in the  $\beta$ -pleated sheet is also quite outstanding.

Although GluDH is known to have two independent binding sites for NADH or NADPH<sup>23</sup> the sequence comparison

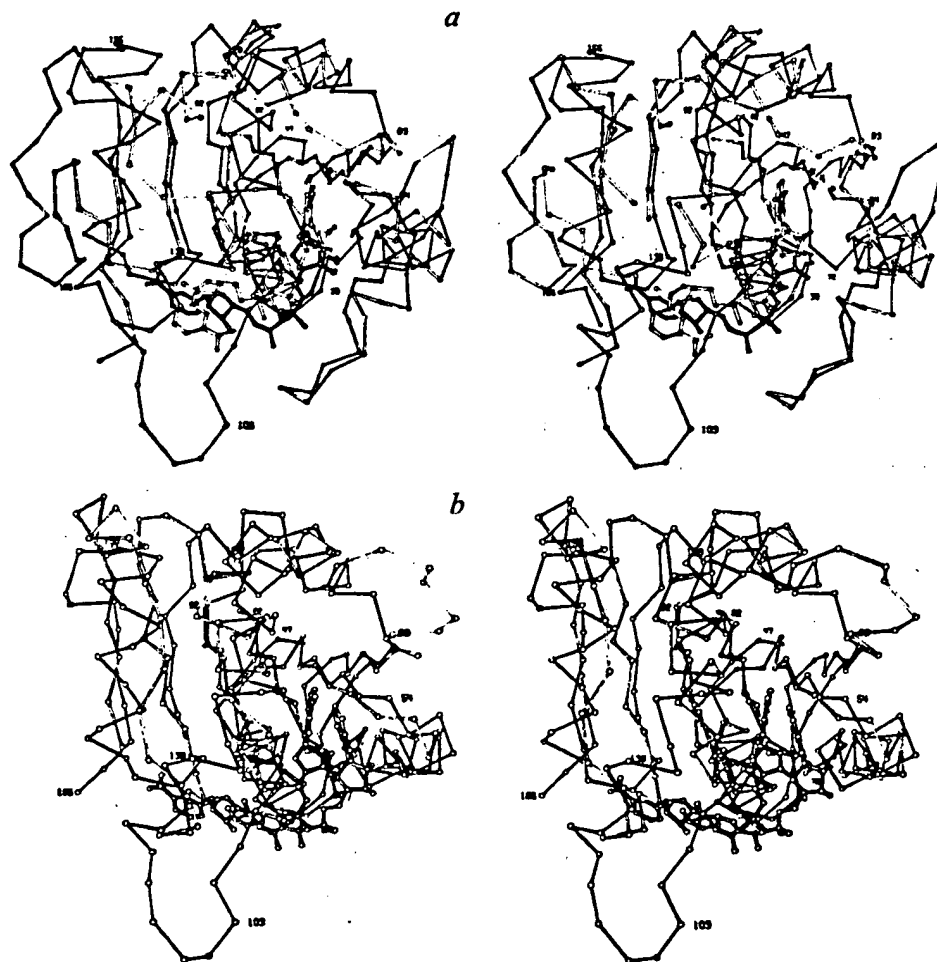


Fig. 3 Superimposed dinucleotide-binding domains using LDH as a standard. The LDH C $\alpha$  backbone is shown in dark as is its NAD coenzyme. Residue numbers are given for LDH only. Corresponding numbers for the other compounds (shown in open bonds) can be derived from Table 1. Comparisons are with (a) GAPDH and (b) LADH.

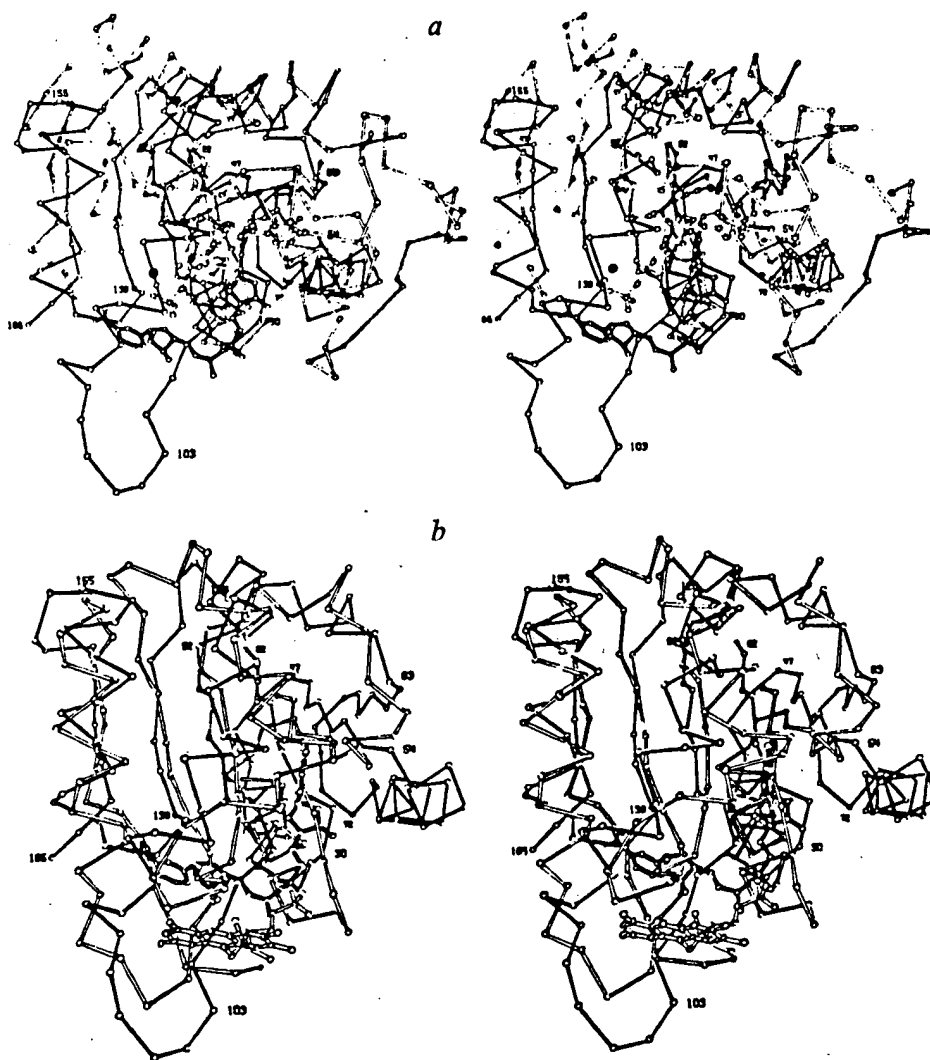


Fig. 4 Comparison of (a) phosphoglycerate kinase (PGK) and (b) flavodoxin with LDH. Conventions used are described in Fig. 3. The black ball is the  $Mg^{2+}$  site in PGK.

could only identify one truly significant portion of the polypeptide chain which compares with the known NAD binding domains in other dehydrogenases.

The alignment shown in Table 1 for GluDH with GAPDH is different from that given by Smith *et al.*<sup>24</sup> or the two comparisons given by Engel<sup>25</sup>, one of which contains Smith's shorter sequence. The alignment given here shows 1.00 minimum base changes per codon with respect to lobster GAPDH as opposed to 1.48 and 1.53 for Engel's relationships. Furthermore, Engel aligns two GluDH sequences with residues 198–261 of GAPDH. However, this region of GAPDH belongs to the domain which supplies catalytic residues and generates substrate specificity, a domain which might be anticipated to be special to GAPDH. Williams and Wilkins<sup>26</sup> point out that Engel's alignments were based on doubtful statistical procedures.

### Measuring evolutionary divergence

A relative estimate of the time since the divergence from a common ancestor (or node in an evolutionary tree) can be obtained by comparing amino acids of related proteins. Margoliash and Fitch<sup>27,28</sup> used the amino acid sequence to show homologous alignments between proteins and then the mean change between the sequence can be measured by various scoring procedures to determine evolutionary distance. In this article (except for the GluDH comparisons) three-dimensional structure has been used to obtain alignments, because of its

greater conservation over amino acid sequence. Every pair of structurally equivalent amino acids can then be examined for their genetic relationship.

It is reasonable to assume that the probability of any one amino acid changing to any other is inversely proportional to the minimum number of base changes. Thus there can be 0, 1, 2 or 3 base changes or on the average 1.5 base changes per codon. If the actual genetic code is taken into account, then the average number of minimum base changes required is 1.57 to change any one amino acid to any other. If, further, the natural frequency of amino acids<sup>29</sup> is considered, this number changes to 1.51. Whether or not an observed minimum base change per codon is significantly lower than a random variation will depend on its distance from the assumed random level. When additional independent data (such as three-dimensional structures) are available, comparisons are of greater significance than when only sequence information is at hand. Thus a combination of structure and sequence permits the measurement of more distant evolutionary relationships.

The minimum base change per codon can be taken as a measure of relative evolutionary distance, that is the elapsed time since the occurrence of a common ancestor for divergent events. This measure is unlikely to be exactly linear with time if no correction is made for back mutations. Another cause for possible deviation from linearity is that the rate of accepting point mutations may have been different at different times. In general, however, a uniform acceptance rate is a reasonable assumption for a given protein. For GAPDH the observed minimum base change per codon between the pig and yeast

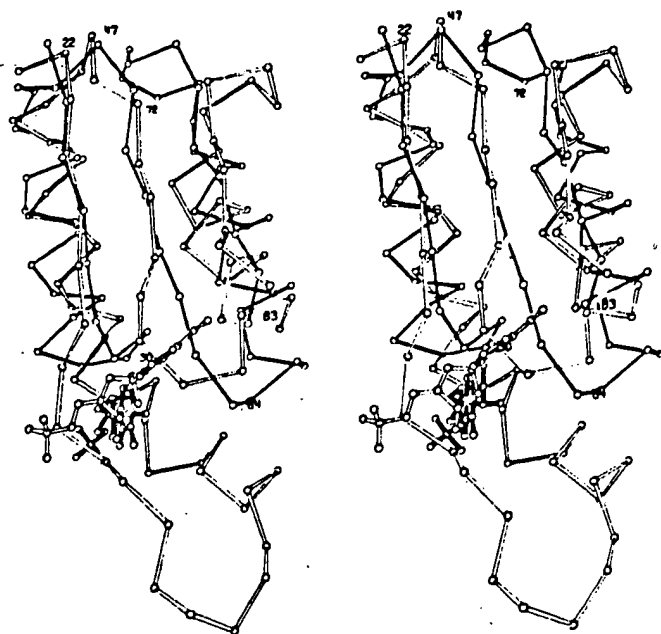


Fig. 5 Superimposed AMP and NMN mononucleotide-binding fragments in LDH. The AMP binding domain is shown with dark bonds and is numbered. The corresponding numbering for the NMN binding fragment can be derived from Table 1.

enzyme in the nucleotide binding protein (residues 1-149) is 0.59, but only 0.35 for residues 150-331. Presumably the differing functions of the two domains within a single polypeptide chain<sup>1</sup> impose different rates of evolution. Thus our estimates of divergence among dehydrogenases have been taken only within the nucleotide-binding domains.

### A time scale for evolution

The evolutionary tree suggested by Buehner *et al.*<sup>8</sup> has been

expanded as shown in Fig. 6. Observed minimum base changes per codon are also related to a probable time at each node. The divergence of the mammalian line from Aves can be placed about  $3 \times 10^8$  yr ago (node 1, Fig. 6), from Arthropoda about  $6 \times 10^8$  yr ago (node 2, Fig. 6) and from Fungi about  $1.2 \times 10^9$  yr ago (node 3, Fig. 6)<sup>30,31</sup>. The presence of the NAD-binding protein in cytoplasmic and mitochondrial dehydrogenases implies that it originated more than  $1.5 \times 10^9$  yr ago when mitochondria were possibly incorporated into proto-eukaryotes<sup>32</sup>. Probably, however, this structure is a good deal older as it would have been required in the earliest prokaryotes for glycolysis around  $3.2 \times 10^9$  yr ago<sup>33</sup>. If it is assumed that simple polypeptides and nucleic acids had gained the ability to perform energy transfer steps and primitive copying processes during precellular evolution<sup>34</sup> we might expect to find the mononucleotide-binding protein before  $3.2 \times 10^9$  yr ago. It is also reasonable to assume that it was formed sometime after the age of the Earth  $4.5 \times 10^9$  yr ago<sup>35</sup>.

Figure 6 shows that, the larger the age associated with the node, the greater is the minimum base change per codon consistent with a divergent evolutionary process from a single common ancestor. That alternative nucleotide-binding proteins exist is, however, evident in the structures of RNase<sup>36</sup> and staphylococcal nuclease<sup>37</sup>. Nevertheless, it might be anticipated that the basic structure shown in Fig. 1 will frequently be found where there is requirement (especially an old and basic requirement) for binding nucleotides. Examples might be amino acid tRNA synthetases, ribosomal proteins and virus coat proteins. The recognition of this structure by sequence homology or from X-ray structure determinations may also give guidance as to function where none is definitely known.

We thank Dr Martha Ludwig, Biophysics Research Division, University of Michigan, Ann Arbor, for information on the hydrogen bonding scheme in *Clostridium* MP flavodoxin and the related amino acid sequence allocations before publication<sup>10</sup>. Similarly we thank Dr Susan Taylor for information on some unpublished segments of the dogfish LDH amino acid sequence and to Drs Colin Blake and Carl Brändén for atomic coordinates of PGK and LADH, respectively. We thank Drs A. Liljas,

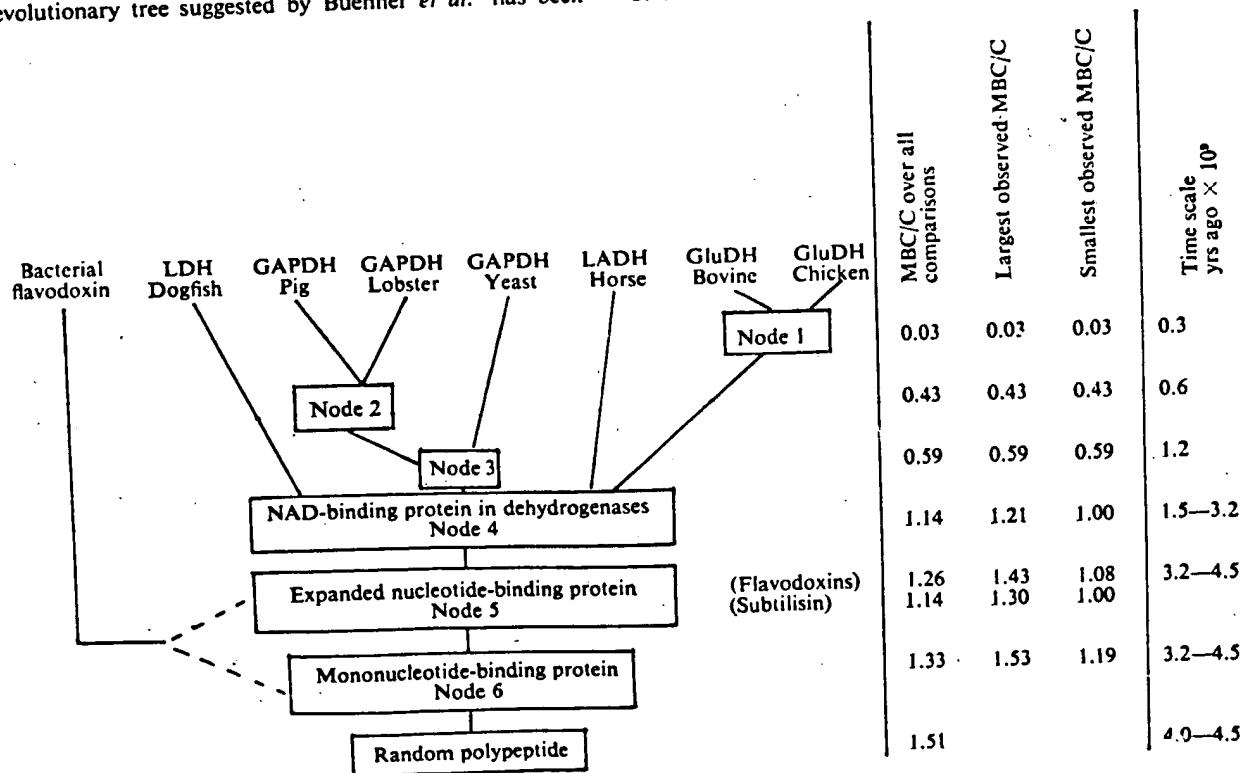


Fig. 6 Evolutionary tree consistent with observed minimum base changes per codon (MBC/C) and possible time scale derived primarily from fossil data.

C. I. Brändén, M. Beuhner, M. L. Hackert and W. Eventoff for important discussions and Sharon Wilder and Sandra Hurto for technical assistance. This work was supported by grants from the US national Science Foundation and the National Institutes of Health. K.W.O. was supported by an American Cancer Society fellowship and D. M. by a CNRS fellowship.

Received December 31, 1973; revised May 20, 1974.

- <sup>1</sup> Rossmann, M. G., and Liljas, A., *J. molec. Biol.* (in the press).
- <sup>2</sup> Adams, M. J., Ford, G. C., Koekoek, R., Lentz, jun., P. J., McPherson, jun., A., Rossmann, M. G., Smiley, I. E., Schevitz, R. W., and Wonacott, A. J., *Nature*, 227, 1098 (1970).
- <sup>3</sup> Rossmann, M. G., Adams, M. J., Buchner, M., Ford, G. C., Hackert, M. L., Lentz, jun., P. J., McPherson, jun., A., Schevitz, R. W., and Smiley, I. E., *Cold Spring Harbour Symp. quant. Biol.*, 36, 179 (1971).
- <sup>4</sup> Hill, E., Tsernoglou, D., Webb, L., and Banaszak, L. J., *J. molec. Biol.*, 72, 577 (1972).
- <sup>5</sup> Brändén, C. I., Eklund, H., Nordström, B., Boiwe, T., Söderlund, G., Zeppezauer, E., Ohlsson, I., and Åkeson, A., *Proc. natn. Acad. Sci. U.S.A.*, 70, 2439 (1973).
- <sup>6</sup> Buehner, M., Ford, G. C., Moras, D., Olsen, K. W., and Rossmann, M. G., *Proc. natn. Acad. Sci. U.S.A.*, 70, 3052 (1973).
- <sup>7</sup> Rao, S. T., and Rossmann, M. G., *J. molec. Biol.*, 76, 241 (1973).
- <sup>8</sup> Watenpugh, K. D., Sieker, L. C., Jensen, L. H., LeGall, T., and Dubourdieu, M., *Proc. natn. Acad. Sci. U.S.A.*, 69, 3185 (1972).
- <sup>9</sup> Andersen, R. D., Apgar, P. A., Burnett, R. M., Darling, G. D., LeQuesne, M. E., Mayhew, S. G., and Ludwig, M. L., *Proc. natn. Acad. Sci. U.S.A.*, 69, 3189 (1972).
- <sup>10</sup> Burnett, R. M., Darling, G. D., Kendall, D. S., LeQuesne, M. E., Mayhew, S. G., Smith, W. W., and Ludwig, M. L., *J. biol. Chem.* (in the press).
- <sup>11</sup> Baltscheffsky, H., *Fourth Int. Conf. on the Origin of Life*, Barcelona, 1973 (in the press).
- <sup>12</sup> Baltscheffsky, H., in *Dynamics of energy transducing membranes* (edit. by Ernster, L., Estabrook, R. W., and Slater, E. C.) (Elsevier, Amsterdam, in the press).
- <sup>13</sup> Blake, C. C. F., and Evans, P. R., *J. molec. Biol.* (in the press).
- <sup>14</sup> Bryant, T. N., Watson, H. C., and Wendell, P. L., *Nature*, 247, 17 (1974).
- <sup>15</sup> Schulz, G. E., and Schirmer, R. H., *Nature* (in the press).
- <sup>16</sup> Smit, J. D. G., thesis, Univ. Groningen (1973).
- <sup>17</sup> Wright, C. S., Alden, R. A., and Kraut, J., *Nature*, 221, 235 (1969).
- <sup>18</sup> Drenth, J., Hol, W. G. J., Jansonius, J. N., and Koekoek, R., *Cold Spring Harbour Symp. quant. Biol.*, 36, 107 (1971).
- <sup>19</sup> Moon, K., Piszkiwicz, D., and Smith, E. L., *Proc. natn. Acad. Sci. U.S.A.*, 69, 1380 (1972).
- <sup>20</sup> Haber, J. E., Koshland, jun., D. E., *J. molec. Biol.*, 50, 617 (1970).
- <sup>21</sup> Barker, W. C., and Dayhoff, M. O., *Atlas of Protein Sequence and Structure*, 5, 101 (1972).
- <sup>22</sup> Jukes, T. H., and Cantor, C. R., in *Mammalian protein metabolism* (edit. by Munro, H. M.), 22 (Academic Press, New York, 1969).
- <sup>23</sup> Krause, J., Bühner, M., and Sund, H., *Eur. J. Biochem.*, 41, 593 (1974).
- <sup>24</sup> Smith, E. L., Landon, M., Piszkiwicz, D., Brattin, W. J., Langley, T. J., and Melamed, M. D., *Proc. natn. Acad. Sci. U.S.A.*, 67, 724 (1970).
- <sup>25</sup> Engel, P. C., *Nature*, 241, 118 (1973).
- <sup>26</sup> Williams, J., and Wilkins, A. G., *Nature*, 247, 556 (1974).
- <sup>27</sup> Margoliash, E., and Smith, E. L., in *Evolving genes and proteins* (edit. by Bryson, V., and Vogel, H. J.), 221 (Academic Press, New York, 1965).
- <sup>28</sup> Margoliash, E., and Fitch, W. M., *Science*, 155, 279 (1967).
- <sup>29</sup> Dayhoff, M. O., Eck, R. V., and Park, C. M., *Atlas of Protein Sequence and Structure*, 5, 89 (1972).
- <sup>30</sup> Dodson, E. O., in *Evolution: Process and product* (Reinhold, New York, 1960).
- <sup>31</sup> Salthe, S. N., in *Evolutionary biology* (Holt, Rinehart and Winston, New York, 1972).
- <sup>32</sup> deDuke, C., *Science*, 182, 85 (1973).
- <sup>33</sup> Barghorn, E. J., *Scientific American*, 224, 30 (1971).
- <sup>34</sup> Carter, C., Kraut, J., *Proc. natn. Acad. Sci. U.S.A.*, 71, 283 (1974).
- <sup>35</sup> York, D., and Farguhar, R. M., in *The Earth's age and geochronology* (Pergamon, Oxford, 1972).
- <sup>36</sup> Richards, F. M., Wyckoff, H. W., Carlson, W. D., Allewell, N. M., Lee, B., and Mitsui, Y., *Cold Spring Harbour Symp. quant. Biol.*, 36, 35 (1971).
- <sup>37</sup> Cotton, F. A., Bier, C. J., Day, V. W., Hazen, jun., E. E., and Larsen, S., *Cold Spring Harbour Symp. quant. Biol.*, 36, 243 (1971).
- <sup>38</sup> Taylor, S. S., Oxley, S. S., Allison, W. S., and Kaplan, N. O., *Proc. natn. Acad. Sci. U.S.A.*, 70, 1970 (1973).
- <sup>39</sup> Harris, J. I., and Perham, R. N., *Nature*, 219, 1025 (1968).
- <sup>40</sup> Davidson, B. E., Sajgö, M., Noller, H. G., and Harris, J. I., *Nature*, 216, 1181 (1967).
- <sup>41</sup> Jones, G. M. T., and Harris, J. I., *FEBS Lett.*, 22, 185 (1972).
- <sup>42</sup> Jörnvall, H., *Eur. J. Biochem.*, 16, 25 (1970).
- <sup>43</sup> Jörnvall, H., *Proc. natn. Acad. Sci. U.S.A.*, 70, 2295 (1973).
- <sup>44</sup> Tanaka, M., Hanice, M., and Yasunabin, K. T., *Biochem. biophys. Res. Commun.*, 44, 886 (1971).
- <sup>45</sup> Fox, J. L., Smith, S. S., and Brown, J. R., *Z. Naturforschung*, 27b, 1096 (1972).
- <sup>46</sup> Alaitan, S. A., DeLange, R. J., and Smith, E. L., *J. biol. Chem.*, 243, 5296 (1968).

## Three abundance classes in HeLa cell messenger RNA

J. O. Bishop, J. G. Morton, M. Rosbash & Melville Richardson

Department of Genetics, University of Edinburgh, West Mains Road, Edinburgh, EH9 3JN, UK

*Approximately 35,000 different poly(A)-containing RNA sequences are present in HeLa cell cytoplasm. The sequences are grouped in three distinct abundance classes.*

THE amount of DNA per haploid genome in the higher eukaryotes is known to be very large, and this has long posed the question of how much of the DNA encodes mRNA. Up to the present time few serious attempts have been made to answer this question. One problem has been the difficulty experienced in attaining sufficiently high concentrations of RNA to drive hybridisation reactions to completion. The discovery of poly(A) stretches in mRNA<sup>1-3</sup> has partially removed this difficulty, since poly(A)-containing RNA can now be separated from rRNA<sup>4,5</sup> and much higher concentrations of mRNA can be attained. A second difficulty has been the existence of repetitive sequences in the DNA<sup>6</sup>: the transcript of one repetitive sequence can hybridise with related sequences,

producing falsely high results. This problem was solved by Hahn and Laird<sup>7</sup> and others<sup>8-11</sup> using a technique devised by Kohne<sup>12</sup>. Labelled nonrepetitive DNA sequences are isolated using hydroxyapatite, and annealed with an excess of unlabelled RNA. The DNA-RNA hybrids are then recovered on hydroxyapatite. In this way it was shown that about 10% of nonrepetitive mouse DNA was represented in transcripts found in total brain RNA<sup>7,9,10</sup>. Similarly 0.9% of nonrepetitive *Xenopus* DNA is represented by transcripts in the mature oocyte<sup>8</sup>.

An even more serious problem has been that whatever answer was obtained, it had to be regarded as a minimum. No matter how much RNA was added, or for how long the samples were annealed, the possibility always remained that classes of RNA were present at a lower concentration than the least concentrated sequences observed to react. Such RNA sequences would fail to react significantly with the complementary DNA sequences, and would go undetected. This problem can now be solved in the following way. It is now possible to synthesise a complementary DNA

## Isolation and Characterization of a Thiamin Pyrophosphokinase Gene, *THI80*, from *Saccharomyces cerevisiae*\*

(Received for publication, February 17, 1993, and in revised form, April 12, 1993)

Kazuto Nosaka‡, Yoshinobu Kaneko§¶, Hiroshi Nishimura, and Akio Iwashima

From the Department of Biochemistry, Kyoto Prefectural University of Medicine, Kamigyo-ku, Kyoto 602 and the §Institute for Fermentation, Osaka, Yodogawa-ku, Osaka 532, Japan.

The *thi80* mutant of *Saccharomyces cerevisiae* (Nishimura, H., Kawasaki, Y., Nosaka, K., Kaneko, Y., and Iwashima, A. (1991) *J. Bacteriol.* 173, 2716-2719) shows markedly reduced activity of thiamin pyrophosphokinase (TPK; EC 2.7.6.2). We have isolated a DNA fragment carrying the *THI80* gene from a yeast genomic library by its ability to complement constitutive synthesis of the thiamin-repressible acid phosphatase, encoded by the *PHO3* gene, of *thi80* mutant cells. On the other hand, the *thi80* locus was found to be located 3.3 centimorgans proximal to the *smg3* locus on the right arm of chromosome XV by genetic mapping analysis, and one more fragment bearing the *THI80* gene trailing *SMP3* gene was obtained by the plasmid eviction method. The nucleotide sequence of the overlapped region between the two isolated DNAs contained an open reading frame of 957 base pairs, encoding a 319-amino acid polypeptide with a calculated molecular weight of 36,616. When the intact *THI80* open reading frame was expressed as a fusion protein carrying three vector-encoded amino acids at its N terminus in *Escherichia coli* lacking TPK, marked TPK activity was detected in the procaryotic cells, proving that the *THI80* gene of *S. cerevisiae* encodes a structural gene of TPK. A gene disruption experiment demonstrated that the *THI80* gene was essential for growth, and therefore, revealed that TPK is the only enzyme capable of synthesizing thiamin pyrophosphate in yeast. Studies of Northern blot analysis and the enzyme assay demonstrated that the *THI80* gene expression is regulated mainly at the mRNA level by the intracellular thiamin pyrophosphate and requires the positive regulatory factors encoded by *THI2* and *THI3* genes. However, unlike thiamin-repressible acid phosphatase and the enzymes involved in thiamin synthesis of *S. cerevisiae*, TPK was found to be expressed constitutively at a low level and incompletely repressed by exogenous thiamin.

Thiamin pyrophosphate (TPP),<sup>1</sup> a coenzyme form of vitamin B<sub>1</sub>, functions in several enzymatic reactions in which TPP serves as a transient intermediate carrier of the aldehyde group (1). Thiamin pyrophosphokinase (TPK; EC 2.7.6.2) catalyzes the direct phosphorylation of thiamin with ATP to form TPP (1). The enzyme has been studied by many investigators and purified from eucaryotic sources, human red blood cells (2), pig heart (3), pig brain (4), parsley leaf (5), and procaryote, *Paracoccus denitrificans* (6). In yeast, TPK was partially purified, and the enzymatic properties and kinetics were well defined (7). However, neither the amino acid nor the nucleotide sequence of TPK has been known in any species.

Recently, we isolated the *thi80* mutant (8) which has a partial deficiency of TPK activity during the study on the regulation of thiamin metabolism in *Saccharomyces cerevisiae*. The *thi80* mutant has the constitutive phenotype of thiamin transport, thiamin-repressible acid phosphatase (T-rAPase) encoded by *PHO3* gene (9), and enzymes involved in thiamin synthesis from 2-methyl-4-amino-5-hydroxymethylpyrimidine and 4-methyl-5- $\beta$ -hydroxyethylthiazole. The mutant also showed high resistance to oxythiamin, a thiamin antagonist whose potency depends on TPK activity (8). In the previous report, we clarified that TPP is a negative effector of the regulatory mechanism of thiamin metabolism in yeast. However, it has not been clear whether the *THI80* gene encodes a structural gene of TPK or a positive regulatory factor specific for expression of a gene encoding for TPK.

We report the isolation, mapping, nucleotide sequence of the *THI80* gene, and the evidence supporting that the *THI80* encodes for the structural gene of TPK by the expression experiment in *Escherichia coli*. We also demonstrated by the gene disruption analysis of *THI80* that TPK is the only enzyme capable of synthesizing TPP in yeast. In addition, the regulation of the *THI80* expression was investigated.

### EXPERIMENTAL PROCEDURES

**Organisms and Cultures**—Table I shows the *S. cerevisiae* strains used in this study. Yeast cells were cultured at 30 °C in YPD medium (1% yeast extract, 2% bacto-peptone, 2% glucose) or in a defined medium containing a 0.67% yeast-nitrogen base (Difco) supplemented with essential amino acids or Wickerham's synthetic minimum medium (12) supplemented with essential amino acids with or without thiamin. Yeast strains auxotrophic for thiamin, *thi2* (13) and *thi3* (11) mutants, were cultured in the minimal medium with thiamin at a concentration of 10<sup>-8</sup> M, which does not cause thiamin repression. The *E. coli* strain MV1184 was used for the amplification of plasmids, and the bacterial cells were grown with shaking at 37 °C in LB broth.

<sup>1</sup> The abbreviations used are: TPP, thiamin pyrophosphate; TPK, thiamin pyrophosphokinase; T-rAPase, thiamin-repressible acid phosphatase; ORF, open reading frame; IPTG, isopropyl-1-thio- $\beta$ -D-galactopyranoside; kb, kilobase pair(s); bp, base pair(s).

\* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) D14417.

‡ To whom correspondence should be addressed: Dept. of Biochemistry, Kyoto Prefectural University of Medicine, Kamigyo-ku, Kyoto 602, Japan. Tel.: 075-251-5315; Fax: 075-213-2746.

¶ Present address: Dept. of Biotechnology, Faculty of Engineering, Osaka University, Suita-shi, Osaka 565, Japan.

TABLE I  
Yeast strains used

Strain	Genotype	Source
X2180-1A	<i>MATa SUC2 mal gal2 CUP1</i>	YGSC <sup>a</sup>
YPH499	<i>MATa ura3-52 his3-Δ200 leu2-Δ1 trp1-Δ63 ade2-101 lys2-801</i>	Y. Ohya <sup>b</sup>
YPH500	<i>MATa ura3-52 his3-Δ200 leu2-Δ1 trp1-Δ63 ade2-101 lys2-801</i>	Y. Ohya <sup>b</sup>
IFO10482	<i>MATa his4-519 gal2</i>	Our stock
T44-5A	<i>MATa thi80-1 ura3-52 leu2-Δ1 trp1-Δ63</i>	Our stock
T48-2D	<i>MATa thi80-1 ura3-52 his3-Δ200 leu2-Δ1 trp1-Δ63</i>	Our stock
T36-2A	<i>MATa thi80-2 pho2-8 pho4-1 trp1 gal2</i>	Our stock
O58-M5	<i>MATa thi2(pho6) gal4</i>	Our stock (10)
TRS3	<i>MATa thi3 leu2-3,112 gal2</i>	Our stock (11)
311/pSMC20-1 <sup>c</sup>	<i>MATa SMP3::URA3 ura3-52 his3-Δ200 leu2-Δ1 trp1-Δ63 ade2-101 lys2-801</i>	This study
YPH501	<i>MATa/MATa ura3-52/ura3-52 leu2-Δ1/leu2-Δ1 his3-Δ200/his3-Δ200 trp1-Δ63/trp1-Δ63 ade2-101/ade2-101 lys2-801/lys2-801</i>	Y. Ohya <sup>b</sup>
NKC2 <sup>d</sup>	<i>MATa/MATa THI80/thi80::URA3 ura3-52/ura3-52 leu2-Δ1/leu2-Δ1 his3-Δ200/his3-Δ200 trp1-Δ63/trp1-Δ63 ade2-101/ade2-101 lys2-801/lys2-801</i>	This study

<sup>a</sup> YGSC is the Yeast Genetic Stock Center, University of California, Berkeley, CA.

<sup>b</sup> University of Tokyo.

<sup>c</sup> Constructed by transformation of YPH499 with a *KpnI* fragment of pSMC20 (representative of 311/pSMC20-1, -2, and -3).

<sup>d</sup> Constructed by transformation of YPH501 with a *SpeI/XhoI* fragment of pAN8-thi80::URA3.

with ampicillin at a concentration of 50 µg/ml. JM105 was used as a host strain for expression of the *THI80* gene.

**Plasmids and DNA Manipulation**—A gene library of *S. cerevisiae* constructed by partial digestion of the genomic DNA with *Sau3AI* and its ligation with YEpl3 (14) at the unique *BamHI* site was obtained from the American Type Culture Collection (Rockville, MD). The shuttle vector pRS316 (15) was utilized for expression in yeast and subcloning of the *THI80* gene, and Ylp5 (16) was used as an integrative vector. A plasmid pSMC20 (a gift of Kenji Irie, Nagoya University) bearing the *SMP3* gene (17) was used for the plasmid eviction method to isolate the *THI80* gene. Fig. 1 shows the structure and construction of the principal plasmids. Yeast chromosomal DNA for Southern blot analysis and plasmid DNA were prepared as described by Hoffman and Winston (18). Yeast transformation was carried out by the method of Ito *et al.* (19). Bacterial transformation and preparation of plasmids and other routine DNA manipulations were performed as described by Sambrook *et al.* (20).

**Genetic Analysis**—Mating, sporulation, dissection, and scoring of genetic markers in yeast and calculation of map distances with tetrad data were carried out by the standard procedures for yeast genetics (21). The T-rAPase activity of *thi80* mutants on plates was detected using the staining method based on the diazo-coupling reaction (22).

**DNA Sequencing**—The DNA sequence was determined on an Automated Laser Fluorescent DNA Sequencer (Pharmacia LKB Biotechnology Inc.) by dideoxy chain termination (23) using an AutoRead Sequencing Kit. In addition to the fluorescently labeled M13 Universal and Reverse primers, further oligonucleotides for walking primer sequencing were synthesized and fluorescently labeled with the Gene Assembler Plus (Pharmacia) according to the manufacturer's manual. The subcloned plasmid pAN7 (Fig. 2) was purified by equilibrium centrifugation in CsCl/ethidium bromide gradients (20) and used as a denatured double-stranded DNA template. The sequencing was performed, twice for each primer, on both orientations. Nucleotide and deduced amino acid sequences were analyzed with programs from a GENETYX software package (Software Development, Tokyo).

**Expression and Preparation of Recombinant Protein**—An expression vector pTrc99A (24) carrying the *trc* promoter, the *lacZ* ribosome-binding site, the *rnnB* transcription terminators, and the multiple cloning site of pUC18 was used. Two oligonucleotide primers were synthesized to amplify the intact *THI80* open reading frame (ORF) by the polymerase chain reaction (Perkin-Elmer Cetus Instruments). The 5'-end primer FT3 is CCGAATTCATGAGCGAGAGTGTATTGA, corresponding to nucleotides 1,428–1,445 with the *EcoRI* linker immediately upstream of the *THI80* ATG, and the 3'-end primer RT1 is GCTTCAGAGGTGACATCCC, corresponding to nucleotides 2,443–2,462 holding *Sall* sites (the nucleotide number refers to Fig. 3B). A plasmid pAN7 (Fig. 2) was used as a template DNA, and conditions of polymerase chain reaction were: 94 °C for 1 min, 50 °C for 1 min, and 72 °C for 2 min. After 30 cycles, the product DNA was purified using Magic™ PCR Preps (Promega). The 1,045-bp product designated HJ-F3R1 containing the *THI80* ORF was

digested with *EcoRI* and *Sall* and subcloned into the *EcoRI* and *Sall* sites of pTrc99A. *E. coli* JM105 harboring pTrc99A-*THI80* was grown at 37 °C to an absorbance of  $A_{600} = 0.8$  in 100 ml culture, and isopropyl-1-thio-β-D-galactopyranoside (IPTG) was added to a final concentration of 0.1 mM. Bacterial cells were further shaken at 28 °C for 10 h. The harvested cells were washed once with 30 mM Tris-HCl (pH 7.5) and 30 mM NaCl and resuspended in 20 ml of the same buffer, followed by disruption (100 watts, 5 min) with an ultrasonic oscillator (Kubota, Tokyo). Cell debris was removed by centrifugation at 20,000 × g for 30 min, and the supernatant was served as a crude extract.

**Gene Disruption**—The 1.1-kb *HindIII* fragment containing yeast *URA3* gene of YEpl24 (25) was inserted into pAN8 (Fig. 2) at the unique *HindIII* site within the *THI80* coding frame to yield pAN8-thi80::URA3. The linear 4.2-kb *SpeI-XhoI* fragment of this plasmid was used for the transformation of diploid yeast YPH501. *Ura*<sup>+</sup> colonies were selected, and gene disruption was confirmed by Southern blot analysis.

**Blot Analysis**—Southern blot analysis was performed as described by Southern (26). Yeast genomic DNA was digested with *EcoRI*, subjected to electrophoresis in a 1% agarose gel, and blotted onto a Hybond-N+ membrane (Amersham). For Northern blot analysis (27), total RNA was isolated from yeast cells by the method of Schneider *et al.* (28). Samples were subjected to electrophoresis in a 1.2% agarose gel containing formaldehyde and blotted onto a Hybond-N+ membrane. Both hybridizations were carried out according to the manufacturer's manual. <sup>32</sup>P-Labeled DNA probes were prepared by the random primer labeling procedure (29) using [ $\alpha$ -<sup>32</sup>P]dCTP (3,000 Ci/mmol) (ICN Biomedicals).

**Enzyme Assays**—TPK activity was measured by the procedure described previously (8). Briefly, the 1.5 ml of assay mixture containing 0.02 M Tris-HCl (pH 8.6), 1 mM thiamin, 1 mM ATP, 2 mM MnSO<sub>4</sub>, and the enzyme source was incubated for 15 min at 37 °C. The reaction was stopped with 0.3 ml of 30% trichloroacetic acid, and after centrifugation the acid was removed twice by extraction with double volumes of ethyl ether. The amount of thiamin and thiamin phosphates in the sample solution were determined by high-performance liquid chromatography after conversion to the corresponding thiochromes (30). Protein was determined using a protein assay kit (Bio-Rad) with bovine serum albumin as the standard.

## RESULTS

**Cloning and Mapping of *THI80***—The *THI80* gene was cloned on the basis of its ability to suppress the *thi80* mutant phenotype producing T-rAPase in the synthetic medium containing  $5 \times 10^{-7}$  M thiamin (8), in which medium yeast phosphate-repressible acid phosphatases are completely repressed (31). A *thi80* mutant, T48-2D (*thi80-1 leu2*), was transformed with the yeast genomic DNA library and plated on the agar medium containing thiamin without leucine. The

plates were incubated at 30 °C for 3 days, and a total of 20,000 *Leu*<sup>+</sup> colonies were examined for the T-rAPase activities by staining. Sixteen colonies did not show T-rAPase activity, and by the plasmid-curing experiment (32) two isolates were confirmed to have a plasmid complementing the *thi80* mutation. These plasmids were prepared from the white (T-rAPase<sup>-</sup>) *Leu*<sup>+</sup> transformants and used to transform *E. coli* MV1184 to Amp<sup>r</sup>. Both plasmids obtained from each Amp<sup>r</sup> transformant conferred the T-rAPase<sup>-</sup> phenotype on yeast strain T48-2D (*thi80-1 leu2*) in minimal medium containing thiamin. In addition, T48-2D bearing these plasmids lost the resistance to oxythiamin at the concentration of 10<sup>-5</sup> M. The obtained two plasmids could be identical based on their restriction maps, and this plasmid, designated pHJ (Fig. 1), had an insertion of a 6.1-kb fragment in YEp13. Transformant carrying pHJ exhibited restoration of the TPK activity and the TPP formation *in vivo* (Table II). To determine whether the pHJ contained the *THI80* gene, we performed site-directed integration of the DNA fragment into the yeast chromosome. A 7.3-kb *NheI-EagI* fragment of pHJ, which carried YEp13-derived DNA at both sides of the inserted 6.1-kb fragment, was subcloned into the *NheI-EagI* gap of Ylp5 to generate pNE80 (Fig. 1). The recombinant plasmid was linearized at the unique *XhoI* site in the cloned fragment and transformed T48-2D (*thi80-1 ura3*). All Ura<sup>+</sup> transformants showed the wild-type phenotype for T-rAPase production. The stable transformant was crossed with the wild-type strain, IFO10482, and the resultant diploid was subjected to tetrad analysis. Of the 12 tetrads tested, all segregated 0+:4- for T-rAPase production on the medium containing thiamin. These findings indicate that pNE80 was integrated at or close to the *thi80* locus and that the cloned DNA carried the *THI80* gene.

On the other hand, when we constructed a *thi80* host strain used in the transformation experiment, we found a significant linkage of *thi80* to *ade2* and *his3* on the right arm of chromosome XV. Tetrad distribution (i.e. parental ditype:nonparental ditype:tetratype) of *ade2* and *thi80* was 47:0:18 and that of *his3* and *thi80* was 42:4:79, indicating that the *thi80* locus is located between the *ade2* and *his3* loci and near the *ade2* locus. Recently, *smc3* mutation (17), which confers stable maintenance of pSR1 (a plasmid of *Zygosaccharomyces rouxii*) in *S. cerevisiae* host, has been mapped between the *ade2* and *his3* loci and close to the *RPB2* locus (34). Then, we examined the linkage of *smc3* and *thi80*. To monitor the segregation of *SMP3* gene by scoring Ura<sup>+</sup> phe-

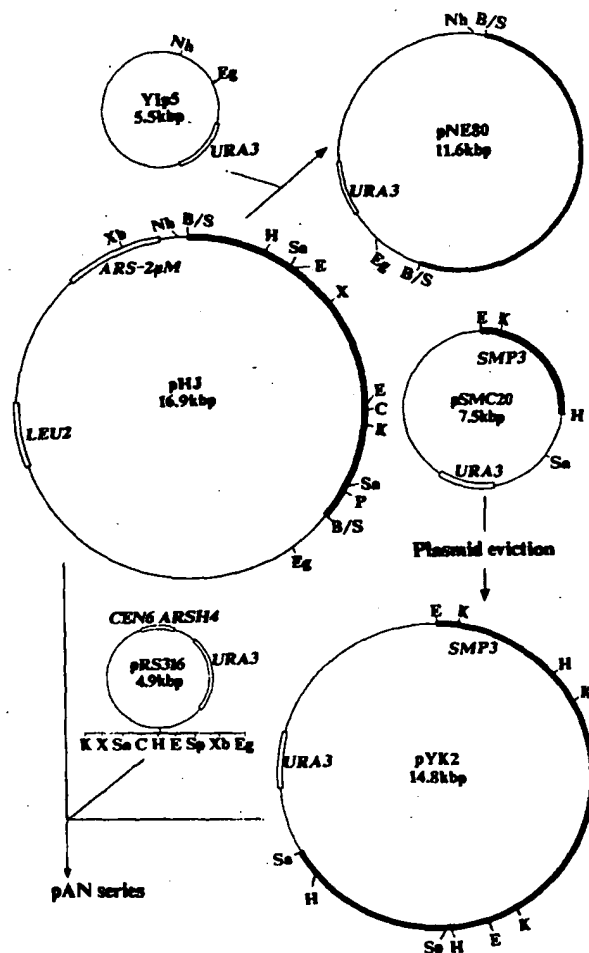


FIG. 1. Structures and constructions of principal plasmids. pSMC20 consists of an *EcoRI-HindIII* fragment containing the *SMP3* gene and the large *EcoRI-HindIII* fragment of Ylp5. Procedures for constructions of pHJ, pYK2, and pNE80 are described in the text. Thick lines indicate the chromosomal DNA of *S. cerevisiae*. Abbreviations of restriction sites: C, *Clai*; E, *EcoRI*; Eg, *EagI*; H, *HindIII*; K, *KpnI*; Nh, *NheI*; P, *PstI*; Sa, *Sall*; Sp, *SpeI*; X, *XhoI*; Xb, *XbaI*. B/S is the junction site of *BamHI* and *Sau3AI*.

notype, we integrated a plasmid pSMC20 into the *SMP3* locus of the *ura3* mutant, YPH499 (*MATa ura3 THI80*), by transformation with linearized pSMC20 DNA at the *KpnI* site located in the *SMP3* coding region. Three independent stable Ura<sup>+</sup> transformants (311/pSMC20-1, -2, and -3) were chosen and crossed to T44-5A (*MATa ura3 thi80-1*). The results of tetrad analysis with the resulting diploids indicated that the *thi80* locus was located 3.3 centimorgans proximal to the *smc3* locus (Table III). Since the relationship of genetic and physical distances of chromosome XV has been reported to be 3.1 kb/centimorgan (36), the *THI80* locus is about 10 kb from the *SMP3* locus. The value of the estimated physical distance moved us to try to clone the *THI80* gene by the plasmid eviction method (37), and one plasmid (pYK2) was rescued from the *Sall*-digested chromosomal DNA prepared from the transformant, 311/pSMC20-1. Then, T48-2D (*thi80-1 ura3*) was transformed by pYK2, and seven out of eight resulting Ura<sup>+</sup> transformants showed wild-type phenotype for T-rAPase production on thiamin containing medium. Therefore, we concluded that pYK2 carried the *THI80* gene. As shown in Fig. 1, the pYK2 contains the *SMP3* gene and an additional 7.9-kb fragment on the opposite side of the *RPB2* locus. A

TABLE II

TPK activity and intracellular concentration of thiamin and thiamin phosphates of *thi80* mutant (T48-2D) and its transformants

TPK activity was determined using the crude extract as a enzyme source obtained by sonic oscillation procedure as described (33) from the yeast cells grown in thiamin-omitted medium. Intracellular contents of thiamin and thiamin phosphates of the cells grown in the medium containing 5 × 10<sup>-7</sup> M thiamin were measured as previously reported (8).

Strain	Plasmid	TPK activity	Intracellular concentration		
			Thiamin	TMP <sup>a</sup>	TPP
		nmol/min/mg protein		μM	
X2180-1A	None	0.27	344	16	208
T48-2D	None	0.07	1497	12	61
	YEp13 <sup>b</sup>	0.08	1385	11	60
	pHJ	1.51	353	27	348

<sup>a</sup> TMP is the abbreviation for thiamin monophosphate.

<sup>b</sup> 2-μm-based multicopy vector used for the construction of the gene library.

### Gene combination and tetrad distribution<sup>a</sup>

Hybrid diploid	Gene combination and tetrad distribution <sup>a</sup>											
	<i>thi80-SMP3::URA3</i>				<i>thi80-ade2</i>				<i>ade2-SMP3::URA3</i>			
	PD	NPD	T	Distance <sup>b</sup>	PD	NPD	T	Distance <sup>b</sup>	PD	NPD	T	Distance <sup>b</sup>
	<i>centi-morgan</i>				<i>centi-morgan</i>				<i>centi-morgan</i>			
311/pSMC20-1xT44-5A	16	0	1		13	0	4		12	0	5	
311/pSMC20-2xT44-5A	13	0	2		11	0	4		10	0	5	
311/pSMC20-3xT44-5A	14	0	0		12	0	3		10	0	4	
Total	43	0	3	3.3	36	0	11	12	32	0	14	15

<sup>a</sup> PD, parental ditype; NPD, nonparental ditype; T, tetratype.

<sup>b</sup> Calculated by the Perkins equation: centimorgan = 100(T + 6NPD)/2(PD + NPD + T) (35).

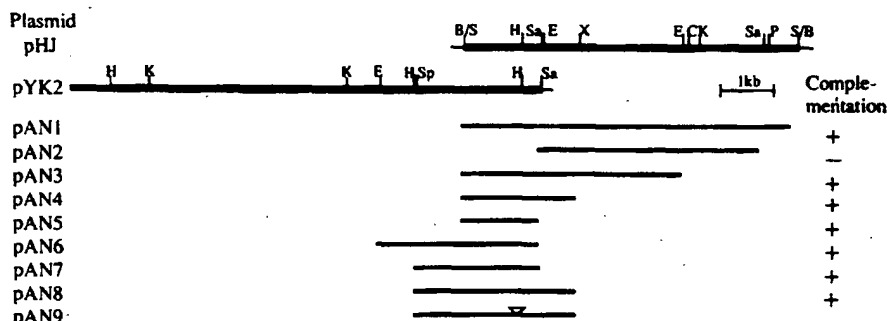


FIG. 2. Restriction maps of *S. cerevisiae* DNAs cloned on pHJ and pYK2 and subcloning of the fragments with the ability to complement the T-rAPase constitutive phenotype of a *thi80* mutant. Thick lines at the top two plasmids indicate the cloned yeast DNA, and the dotted line of pYK2 shows the DNA fragment of *SMP3* gene. Thin lines of pAN plasmids are the subcloned fragments on the plasmid, pRS316, and each insert has a multicloning site at each end. The triangle on the fragment of pAN9 indicates the site modified on cleavage with *Hind*III, followed by treatment with Klenow fragment and religation. The ability of each plasmid to complement the *thi80* mutant phenotype is indicated in the right-hand column: +, complementation; -, no complementation. Abbreviations of restriction sites are the same as in the legend to Fig. 1.

homolog of THI80 in the NBRF database, indicating that the THI80 protein is a new protein.

**Expression of *THI80* ORF in *E. coli***—To ascertain whether the *THI80* gene encodes the structural gene of TPK, we attempted to express the *THI80* protein in *E. coli*, in which TPK dose not exist, and TPP is synthesized from thiamin monophosphate by thiamin-monophosphate kinase (EC 2.7.4.16) (40). The polymerase chain reaction product (HJ-F3R1) containing the complete *THI80* ORF with *Eco*RI site immediately upstream of the *THI80* ATG and *Sal*I site located in the 3'-noncoding region was prepared and ligated between these restriction sites into expression vector, pTrc99A. The resultant translational initiation sequence is as follows.

Met Glu Phe Met Ser Glu  
5'...AGGAAACAGACC ATG GAA TTC ATG AGC GAG...3'  
EcoRI  
(Sequence 1)

The expected fusion protein has an  $M_r$  of 37,023 and carries three vector-encoded amino acids (Met, Glu, Phe) at its N terminus followed by 319 TH180-specific amino acids. Bacterial cells harboring the plasmid were cultured at 28 °C with or without 0.1 mM IPTG induction, in which condition the fusion protein could be obtained in the soluble fraction of cell extract without accumulation of protein aggregates. Fig. 4A shows the patterns of SDS-polyacrylamide gel electrophoresis of the crude extracts of the two independent pTcr99A-TH180 transformants. A 36-kDa band, which was almost consistent with the expected size of the fusion protein, appeared in both IPTG-induced cells (*lanes 3 and 4*).

Then the TPK activities in the cell-free extract were determined after the samples were dialyzed sufficiently. Fig. 4B

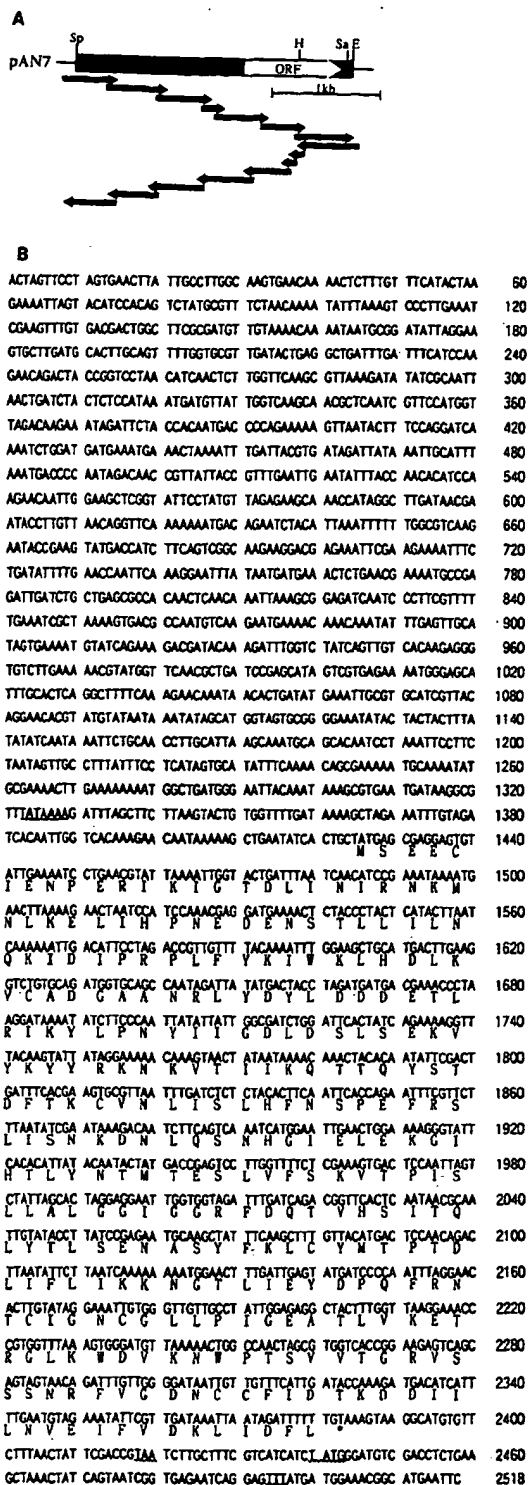
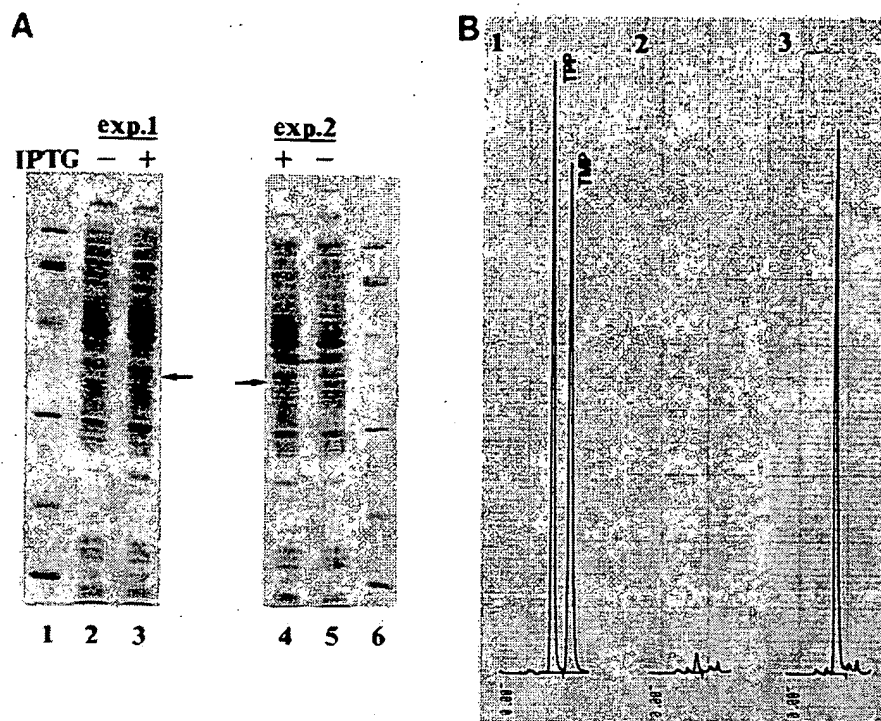


FIG. 3. Sequencing of *THI80*. **A**, sequencing strategy. Open box indicates the position and direction of the *THI80* ORF. Arrows show the direction and approximate extent of sequence determination by primer walking. Abbreviations used are the same as in the legend to Fig. 1. **B**, *THI80* sequence. The predicted amino acid sequence is shown in a one-letter code below the nucleotide sequence. The putative TATA box is underlined, and the wavy underlines indicate possible signals for transcription termination.

shows the chromatogram of thiocromes from thiamin phosphates in the assay mixture for TPK activity. Marked TPK activity was detected in the IPTG-induced cells (i.e. 43.2 nmol of TPP produced per min/mg of protein), whereas no enzyme activity was detected in the uninduced cells (panels 2 and 3). When thiamin monophosphate, toward which the TPK of yeast is completely inactive (1), was used as substrate in place of thiamin, TPP formation was not observed in the IPTG-induced cells (data not shown). Moreover, potassium ions, which are required for the thiamin-monophosphate kinase activity of *E. coli* (40), were not contained in the mixture used for TPK activity. These results clearly indicated that TPK were produced in the IPTG-induced cells carrying pTrc99A-*THI80* and that the *THI80* gene of *S. cerevisiae* encodes a structural gene of TPK.

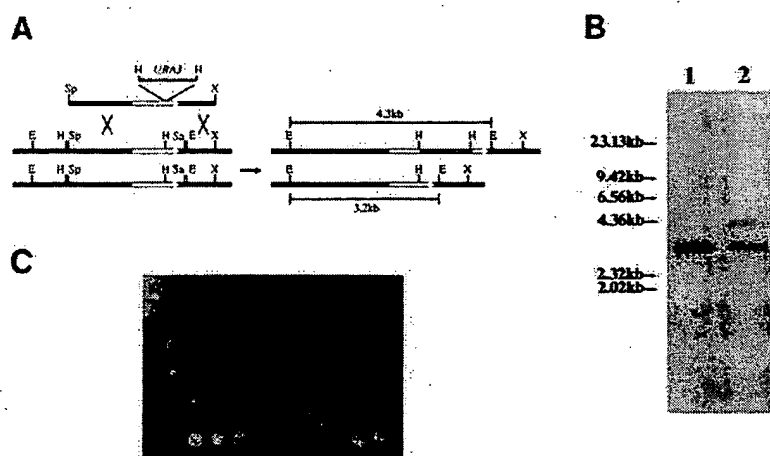
**Disruption of *THI80***—We disrupted the corresponding gene locus to confirm the identity of the cloned gene further. The *URA3* fragment was inserted into pAN8 at the unique *HindIII* site within the *THI80* coding region to create pAN8-*thi80::URA3*. The disrupted *thi80* fragment was excised from the plasmid and introduced into *ura3/ura3* diploid YPH501 (Fig. 5A). DNA was extracted from a *Ura*<sup>+</sup> transformant (NKC2) and simultaneously digested with *EcoRI*, the resulting fragments were separated by agarose gel electrophoresis, and Southern blot analysis was carried out using HJ-F3R1 as a probe. Fig. 5B indicates that the *THI80* gene has been correctly disrupted. NKC2 was a heterozygous diploid carrying the wild-type and disrupted alleles of *THI80* (lane 2), whereas the parent YPH501 contained only the wild-type allele (lane 1). When NKC2 was dissected, less than two spore clones survived as shown in Fig. 5C, and all the spore clones were *Ura*<sup>-</sup>. This finding demonstrated that the *THI80* is an essential gene for growth, and therefore, proved that TPK is the only enzyme capable of synthesizing TPP in *S. cerevisiae*.

**Regulation of *THI80* Expression**—We have already reported that the activities of T-rAPase and the enzymes involved in thiamin biosynthesis in *S. cerevisiae* are coordinately repressed by exogenous thiamin (13). We also elucidated that expression of these genes is controlled by two positive regulatory genes, *THI2* and *THI3*, which might be negatively regulated by the intracellular TPP level via some gene(s) (8, 11). To determine whether expression of the *THI80* gene is regulated in a similar manner to the enzymes involved in thiamin metabolism, the effect of thiamin in the medium on the abundance of *THI80* mRNA and TPK activity was investigated using the wild-type strain, *thi2*, *thi3*, and *thi80* mutants. Cells were cultured under either low or high concentration of thiamin, total RNA was extracted, and then Northern blot analysis was performed using *THI80*, *PHO3*, and *URA3* DNAs as probes, and simultaneously the TPK activity in the soluble fraction from the cells was measured. *THI80* mRNA was detected as a single 1.2-kb band (Fig. 6A, upper panel), which was almost consistent with the coding region. The abundance of *THI80* mRNA in the wild-type strain was reduced certainly by thiamin in the medium (lane 1, 2), whereas the *thi2* and *thi3* mutants showed low levels of *THI80* mRNA (lanes 3–6). These levels of *THI80* mRNA were comparable with those of the TPK activity (Fig. 6B), indicating that *THI80* expression in the wild-type strain, like the *PHO3* gene, is controlled at the mRNA level by thiamin in the medium and requires the positive factors, *THI2* and *THI3*. In *thi80-2* mutant which showed the TPK activity of about one tenth of that in the wild-type strain (Fig. 6B), the *THI80* mRNA was expressed at the same level as that of the wild-type strain (Fig. 6A, lane 1, 7), suggesting that the mutation site of *thi80-2* is in the coding region of the mRNA to disturb



**FIG. 4. Expression of *THI80* in *E. coli*.** The growth condition of the bacterial cells bearing the expression vector pTrc99A-*THI80* and preparation of crude extracts are described under "Experimental Procedures." **A**, SDS-polyacrylamide gel electrophoresis. Crude extracts (18  $\mu$ g of protein) from two independent transformants grown in uninduced (lanes 2 and 5) and IPTG-induced (lanes 3 and 4) cultures were analyzed by 12% SDS-polyacrylamide gel electrophoresis using the method of Laemmli (41) followed by Coomassie Blue staining. The position of presumed fusion *THI80* protein is indicated by the arrow. Lanes 1 and 6 show molecular standard proteins (Bio-Rad): phosphorylase *b* (97,400), bovine serum albumin (66,200), ovalbumin (45,000), carbonic anhydrase (31,000), soybean trypsin inhibitor (21,500), lysozyme (14,400). **B**, chromatogram of thiochromes from thiamin phosphates on high-performance liquid chromatography. Crude extracts (10  $\mu$ g of protein) from the uninduced (panel 2) and IPTG-induced (panel 3) cells were incubated for 15 min at 37 °C in the TPK assay mixture, and 100  $\mu$ l of the solution after conversion to the thiochromes as described under "Experimental Procedures" were injected onto the column. Panel 1 indicates the chromatogram from authentic thiamin phosphates (100 pmol for each chemicals). *TMP* is the abbreviation for thiamin monophosphate.

**FIG. 5. Disruption of *THI80*.** **A**, disruption strategy. The procedure is described under "Experimental Procedures." The white bar with an arrow indicates the coding region. Abbreviations of restriction sites are the same as in the legend to Fig. 1. **B**, confirmation of the expected DNA change by Southern analysis. Lane 1, parental diploid YPH501; lane 2, transformed diploid NKC2. A *Hind*III digest of  $\lambda$ -phage DNA was used as size markers and are shown on the left side of the figure. **C**, tetrad analysis of NKC2. Dissection agar slab containing tetrad sets was placed on YPD and photographed after incubation at 30 °C for 3 days.



the stability or function of the TPK polypeptide. Furthermore, the decrease in the *THI80* mRNA level by thiamin in *thi2*, *thi3* and *thi80-2* mutants was low (lane 4, 6, 8), which was thought to be caused by an insufficient increase in the intracellular concentration of the corepressor, TPP (8), due to the low levels of TPK activity in these mutants (Fig. 6B). These findings suggested that the *THI80* gene expression is regulated negatively by TPP, the product synthesized by TPK. However, there was a significant difference in the extent of repression by thiamin between *PHO3* and *THI80* genes in

the wild-type strain (Fig. 6A, lane 1, 2). Thiamin in the medium repressed almost completely *PHO3* mRNA in the wild-type strain, whereas the decrease in the *THI80* mRNA level by thiamin was not as marked as that of *PHO3* mRNA. In addition, TPK activity was no longer decreased when the wild-type strain was grown in the medium containing  $10^{-6}$  M thiamin (data not shown). These findings suggested that the expression of the *THI80* gene is incompletely repressed by thiamin and the basal level expression is not under the control of the thiamin regulatory system.

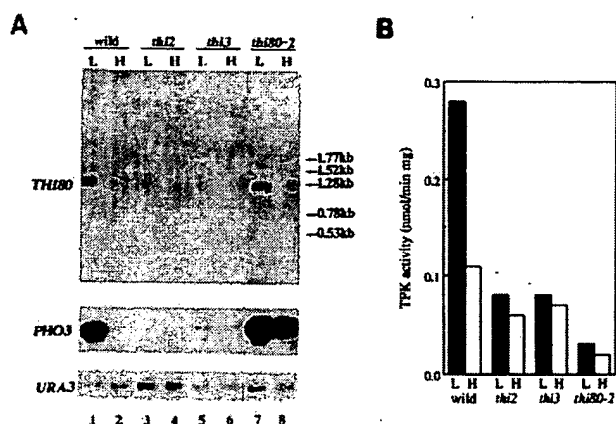


FIG. 6. Regulation of *THI80*. Yeast strains X2180-1A (wild type), O58-M5 (*thi2*), TRS3 (*thi3*), and T36-2A (*thi80-2*) were grown in minimal medium containing  $10^{-8}$  (L) or  $10^{-6}$  (H) M thiamin. A, Northern blot analysis. Total RNA was isolated and subjected to Northern blot analysis (10  $\mu$ g/lane) as described under "Experimental Procedures." Upper panel, the  $^{32}$ P-labeled HJ-F3R1 containing intact *THI80* ORF was used as probe. RNA molecular weight standards (0.16–1.77-kb RNA Ladder, Life Technologies, Inc.) are indicated on the right. Middle panel, the 3.2-kb *Pst*I-*Bam*HI fragment containing *PHO3* gene was prepared from plasmid pKB (42) and used as probe. Lower panel, the 1.1-kb *Hind*III fragment of *URA3* from plasmid YEp24 (25) was used as probe. B, enzyme assay. Crude extracts were prepared by sonication of yeast cells (33), and TPK activity was determined as described under "Experimental Procedures."

#### DISCUSSION

We have obtained two DNA fragments to complement the constitutive phenotype of T-rAPase of the *thi80* mutant: one from a yeast genomic library and the other by the plasmid eviction method. The site-directed integration analysis of the DNA fragment from the library revealed that the cloned DNA carries the *THI80* gene and not a suppressor gene of *thi80*. Our cloned fragment from the library hybridized in a yeast chromosome blot to chromosome XV (data not shown). This finding shows an agreement with the genetic mapping analysis which demonstrated that the *thi80* locus is located 3.3 centimorgans proximal to the *smp3* locus on the right arm of chromosome XV. On the other hand, the other DNA fragment obtained by the plasmid eviction method displayed that the about 8-kb downstream regions from the *SMP3* ORF were similar in the restriction map to the fragment from the gene library, indicating that the latter fragment also contained the *THI80* gene and that the *THI80* ORF is located within the overlapped region of the two cloned fragments.

The *THI80* gene revealed an ORF of 957 nucleotides encoding a protein of 319 amino acids with a predicted molecular weight of 36,616. This is in good agreement with the identification of a 1.2-kb transcript by Northern hybridization analysis. The expression experiment of the intact *THI80* ORF in *E. coli* indicates that the *THI80* gene of *S. cerevisiae* is a structural gene of TPK. When the synthesized DNA corresponding to nucleotides 1,498–1,517 with *Eco*RI linker was used as the 5'-end primer in place of the FT3 primer in polymerase chain reaction with 3'-end primer RT1, and the resultant product DNA was ligated in expression vector pTcr99A, which was designed so that the second methionine of intact *THI80* protein follows the vector-encoded amino acids, no TPK activity was detected in the extract of the transformed *E. coli* cells induced by IPTG (data not shown). We, therefore, considered that the first methionine of the *THI80* ORF is the initiator methionine, and the N-terminal sequence is essential for enzyme activity. We cannot find a

homolog of the amino acid sequence predicted from the *THI80* ORF in the database. However, the sequence starting at the 11th residue, Arg-X-X-Gly-X-X-Leu-Ile-Asn-Ile-, of the *THI80* protein resembles the ATP binding motif observed in  $\alpha$  and  $\beta$  subunits of ATP synthase and other adenine nucleotide binding proteins (43). The reason why the truncated *THI80* protein which lacks the first 24 residues expressed in the bacterial cells has no TPK activity is attributed to the loss of the presumed ATP-binding domain, although studies by means of site-directed mutagenesis and protein engineering are required.

Purified TPK has been prepared from pig heart (3), pig brain (4), parsley leaf (5), and *P. denitrificans* (6). Recently, Egi *et al.* reported the isolation and purification of TPK from human red blood cells (2). All investigated TPKs had a dimeric structure expected for that from the plant. From the comparison of the elution volume of TPK activity with those of standard proteins by gel filtration analysis using Sephacryl S-200, the molecular weight of the fusion protein was estimated to be 72,000 (data not shown). These findings suggest that the native TPK of *S. cerevisiae* exists in the dimer form of identical subunits. Hamada (3) indicated that pig heart TPK forms disulfide cross-linkages between two identical subunits. The predicted protein encoded by the *THI80* ORF contains 8 cysteine residues, some of which may be involved in the linkage between the *THI80* polypeptides.

The biosynthesis of TPP in *S. cerevisiae* involves the independent formation of the pyrimidine and thiazole moiety of thiamin, their condensation to thiamin monophosphate, and the subsequent hydrolysis of thiamin monophosphate and pyrophosphorylation of thiamin formed (1). Although free thiamin is certainly thought to be an obligate intermediate between thiamin monophosphate and TPP, Tokuda (44) suggested the existence of the enzyme-catalyzing direct phosphorylation of thiamin monophosphate to form TPP in yeast. However, the gene disruption experiment in this study plainly indicated that TPP can be synthesized by nothing but TPK in *S. cerevisiae*. Previous studies from our laboratory have shown that thiamin phosphates are hydrolyzed by T-rAPase in the periplasmic space into free thiamin which can be taken up by an active transport system in *S. cerevisiae* (30, 45). To confirm that yeast cells have no ability to transport TPP in phosphate form, we planted the spores dissected from the diploid, NKC2, on YPD plates containing  $10^{-4}$  M TPP instead of ordinary YPD medium in the course of the *THI80* disruption experiment. The result for eight tetrads tested was the same as that using the medium without TPP (data not shown), indicating that TPP itself cannot be taken up by yeast cells and the *THI80* is unconditionally an essential gene in *S. cerevisiae*.

The findings obtained by Northern blot hybridization analysis and enzyme assay suggested that the *THI80* gene expression of *S. cerevisiae* is regulated mainly at the mRNA level by a process that requires the positive factors, *THI2* and *THI3*, and the negative factor, TPP. However, the TPK seems to be expressed constitutively at a low level and incompletely repressed by exogenous thiamin, whereas the enzymes involved in thiamin biosynthesis, T-rAPase encoded by *PHO3*, and the thiamin transport system are almost completely repressed by thiamin (13). These findings were conceivable, because the enzymatic formation of TPP by TPK encoded by *THI80* is the only process for supplying this coenzyme from thiamin in *S. cerevisiae*. Recently, we localized the region responsible for the transcriptional activation of the *PHO3* gene in response to thiamin in the medium (42). Although comparison and best alignment of the 400-bp 5'-upstream

noncoding region of the *PHO3* and *THI80* genes showed approximately 52% homology, we have not yet determined the consensus sequence concerned in thiamin regulation between these two genes. It is also unknown at present whether a change in *THI80* mRNA abundance by intracellular TPP concentration in yeast cells is due to the regulation at the transcription of the *THI80* gene or the degradation of the mRNA. To clarify the detailed mechanism of the regulation of *THI80* expression, further studies at the molecular level are required.

## REFERENCES

- Leder, I. G. (1975) in *Metabolic Pathway: Metabolism of Sulfur Compounds* (Greenberg, D. M., ed) 3rd Ed., Vol. 7, pp. 57-85, Academic Press, New York
- Egi, Y., Koyama, S., Shioda, T., Yamada, K., and Kawasaki, T. (1992) *Biochim. Biophys. Acta* 1180, 171-178
- Hamada, M. (1969) *Seikagaku* 41, 837-849
- Wakabayashi, Y. (1978) *Vitamins (Kyoto)* 52, 223-236
- Mitsuda, H., Takii, Y., Iwai, K., and Yasumoto, K. (1975) *J. Nutr. Sci. Vitaminol.* 21, 103-115
- Sanemori, H., and Kawasaki, T. (1980) *J. Biochem. (Tokyo)* 88, 223-230
- Kajiro, Y. (1959) *J. Biochem. (Tokyo)* 46, 1523-1539
- Nishimura, H., Kawasaki, Y., Nosaka, K., Kaneko, Y., and Iwashima, A. (1991) *J. Bacteriol.* 173, 2716-2719
- Schweingruber, M. E., Fluri, R., Maundrell, K., Schweingruber, A.-M., and Dummermuth, E. (1986) *J. Biol. Chem.* 261, 15877-15882
- Toh-e, A., Kakimoto, S., and Oshima, Y. (1975) *Mol. & Gen. Genet.* 141, 81-83
- Nishimura, H., Kawasaki, Y., Kaneko, Y., Nosaka, K., and Iwashima, A. (1992) *J. Bacteriol.* 174, 4701-4706
- Wickerham, L. J. (1951) *U. S. Dep. Agric. Tech. Bull.* 1029, 1-56
- Kawasaki, Y., Nosaka, K., Kaneko, Y., Nishimura, H., and Iwashima, A. (1990) *J. Bacteriol.* 172, 6145-6147
- Broach, J. R., Strathern, J. N., and Hicks, J. B. (1979) *Gene (Amst.)* 8, 121-133
- Sikorski, R. S., and Hieter, P. (1989) *Genetics* 122, 19-27
- Struhl, K., Stinchcomb, D. T., Scherer, S., and Davis, R. W. (1979) *Proc. Natl. Acad. Sci. U. S. A.* 76, 1035-1039
- Irie, K., Araki, H., and Oshima, Y. (1991) *Mol. & Gen. Genet.* 225, 257-265
- Hoffman, C. S., and Winston, F. (1987) *Gene (Amst.)* 57, 268-272
- Ito, H., Fukuda, Y., Murata, K., and Kimura, A. (1983) *J. Bacteriol.* 153, 163-168
- Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Rose, M. D., Winston, F., and Hieter, P. (1990) *Methods in Yeast Genetics: A Laboratory Course Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Hansche, P. E., Beres, V., and Lange, P. (1978) *Genetics* 88, 673-687
- Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* 74, 5463-5467
- Amann, E., Ochs, B., and Abel, K.-J. (1988) *Gene (Amst.)* 69, 301-315
- Botstein, D., Falco, S. C., Stewart, S. E., Brennan, M., Scherer, S., Stinchcomb, D. T., Struhl, K., and Davis, R. W. (1979) *Gene (Amst.)* 8, 17-24
- Southern, E. M. (1975) *J. Mol. Biol.* 98, 503-517
- Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. U. S. A.* 77, 5201-5205
- Schneider, R., Gander, I., Müller, U., Mertz, R., and Winnacker, E. L. (1986) *Nucleic Acids Res.* 14, 1303-1317
- Feinberg, A. P., and Vogelstein, B. (1984) *Anal. Biochem.* 137, 266-267
- Nosaka, K., Kaneko, Y., Nishimura, H., and Iwashima, A. (1989) *FEMS Microbiol. Lett.* 60, 55-60
- Lemire, J. M., Willcock, T., Halvorson, H. O., and Bostian, K. A. (1985) *Mol. Cell. Biol.* 5, 2131-2141
- Rose, M. D., and Broach, J. R. (1991) *Methods Enzymol.* 194, 195-230
- Iwashima, A., Nishimura, H., and Nose, Y. (1979) *Biochim. Biophys. Acta* 557, 460-468
- Sweetser, D., Nonet, M., and Young, R. A. (1987) *Proc. Natl. Acad. Sci. U. S. A.* 84, 1192-1196
- Perkins, D. D. (1949) *Genetics* 34, 607-626
- Mortimer, R. K., Schild, D., Contopoulou, C. R., and Kans, J. A. (1989) *Yeast* 5, 321-404
- Rothstein, R. (1991) *Methods Enzymol.* 194, 281-301
- Zaret, K. S., and Sherman, F. (1982) *Cell* 28, 563-573
- Langford, C. J., and Gallwitz, D. (1983) *Cell* 33, 519-527
- Nishino, H., Iwashima, A., and Nose, Y. (1971) *Biochem. Biophys. Res. Commun.* 45, 363-368
- Laemmli, U. K. (1970) *Nature* 227, 680-685
- Nosaka, K., Yamanishi, K., Nishimura, H., and Iwashima, A. (1992) *FEBS Lett.* 306, 244-248
- Walker, J. E., Saraste, M., and Gay, N. J. (1984) *Biochim. Biophys. Acta* 788, 184-200
- Tokuda, Y. (1964) *Vitamins (Kyoto)* 29, 40-46
- Nosaka, K. (1990) *Biochim. Biophys. Acta* 1037, 147-154